

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE INFORMÁTICA**

**MÁSTER EN INTERNET DE LAS COSAS**



**TRABAJO FIN DE MÁSTER  
CURSO 2017-2018**

**VISUALIZACIÓN DE DATOS Y PERSONALIZACIÓN  
INTELIGENTE DE EXPLICACIONES PARA SU  
UTILIZACIÓN EN LA MEDICINA DEL DEPORTE**

**Autor:**

Jaime Nicolás García Giménez

**Directores:**

María Belén Díaz Agudo

Juan Antonio Recio García

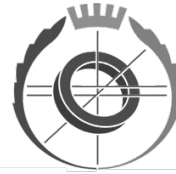
**Convocatoria: Septiembre 2018**

**Calificación: Notable 8**





UNIVERSIDAD COMPLUTENSE  
MADRID



**Facultad de Informática**

D.<sup>a</sup> MARÍA BELÉN DÍAZ AGUDO Y D. JUAN ANTONIO RECIO GARCÍA,  
DOCTORES DEL GRUPO DE APLICACIONES DE INTELIGENCIA  
ARTIFICIAL DE LA FACULTAD DE INFORMÁTICA DE LA  
UNIVERSIDAD COMPLUTENSE DE MADRID.

CERTIFICAN: Que la presente memoria titulada “*VISUALIZACIÓN Y PERSONALIZACIÓN DE SISTEMAS INTELIGENTES DE ELECTROMIOGRAFÍA PARA SU UTILIZACIÓN EN LA MEDICINA DEL DEPORTE*” constituye la Memoria de Trabajo Fin de Máster presentada como requisito para optar al título de Máster en Internet de las Cosas por la Universidad Complutense de Madrid, y ha sido realizada bajo nuestra tutela.

Y para que así conste, firman el presente certificado en Madrid a 20 de septiembre de 2018.

Fdo.: María Belén Díaz Agudo

Fdo.: Juan Antonio Recio García



## **Autorización de difusión y utilización**

El autor del presente Trabajo Final de Máster firma la autorización de difusión y utilización del material académico que en este mismo se incluye, cuyo desarrollo se encuentra dentro del marco de los proyectos TIN2014-55006-R y TIN2017-87330-R.

En Madrid a 20 de septiembre de 2018.

Fdo.: Jaime Nicolás García Giménez



## Resumen

El presente documento refleja el Trabajo Final correspondiente al Máster Universitario en Internet de las Cosas, y se ubica en el área de aplicación de sistemas inteligentes. En el desarrollo del proyecto se tratan generalmente los principios de visualización aplicables en cualquier entorno del Internet de las Cosas y que pueden estar dirigidos a equipos y usuarios de todo tipo.

Hoy en día se tratan métodos de recogida y procesamiento de datos altamente heterogéneos, orientando la capacidad de conocimiento de esos datos a una cantidad de usuarios también muy heterogénea. De este modo, la generación de explicaciones visuales e interactivas dentro de plantillas de informes se puede optimizar mediante el uso de técnicas de Inteligencia Artificial (IA) como el Case-Based Reasoning (CBR), capaz de abordarse junto al diseño e implementación de herramientas de visualización avanzada.

Este conjunto de aportaciones inteligentes se traduce en una personalización total de los informes, y posee la finalidad última de facilitar la cognición y detección de relaciones entre patrones dentro de los casos tratados mediante técnicas avanzadas de visualización de datos, además de capacitar la automatización de decisiones a partir de la observación y evaluación de los datos asociados a dichos casos. Así pues, resulta de especial interés el hecho de trasladar este objetivo a la Medicina y aplicar técnicas de IA como el CBR en sistemas compuestos por sensores de electromiografía (EMG), con tal de explicar a los expertos las características de los datos de pacientes y así poder facilitar a la toma de decisiones en el día a día.

Este Trabajo Final de Máster conforma una división importante dentro de un ambicioso proyecto de investigación, y el objetivo primordial definido para su ejecución se basa en el desarrollo de un prototipo de mecanismo inteligente de generación de explicaciones visuales personalizables e interactivas con capacidad de aprendizaje a partir de la interacción. Aunque el mecanismo y la metodología se establecen de forma genérica, la prueba de concepto se realiza con datos correspondientes al dominio de la Medicina del Deporte, concretamente proporcionando soporte al profesional médico en el proceso de interpretación de características de datos asociados a pacientes que se corresponden con análisis de electromiografía.

Para ello, en primer lugar, se describen los procesos y jerarquías CBR en la reutilización de elementos visuales en informes. Seguidamente, se analiza la naturaleza de las técnicas de visualización de datos de una forma genérica, con tal de a posteriori focalizar en un perfil de entidad concreto dentro del campo que ocupa la Medicina del Deporte. Este enfoque se aborda a lo largo del proyecto a modo de estudio práctico, y en todo momento se persigue explotar la funcionalidad de la herramienta perteneciente a dicha entidad para seguidamente potenciar su evolución tanto desde la perspectiva descriptiva como predictiva. Este análisis de datos médicos es causante de que, en última instancia, se estudien los procedimientos legislativos para el tratamiento de los propios datos dentro del ámbito sanitario.

**Palabras clave:** *Visualización, CBR, IA, Big Data, Medicina del Deporte, EMG, storytelling, interacción, explicaciones, análisis, conocimiento, reutilización, adaptación, gráficos, automatización, legalidad.*

## **Abstract**

This document reflects the Final Master Project corresponding to the Master's Degree in Internet of Things, developed within the area of intelligent systems' application. During the progress of the project, the visualization principles applicable in any environment of the Internet of Things are explained generally and aimed to any kind of equipment and final user.

Nowadays, highly heterogeneous data collection and processing methods are used, orienting the knowledge capacity of these data to a very heterogeneous number of users. In this way, the generation of visual and interactive explanations within report templates can be optimized through the use of Artificial Intelligence (AI) techniques such as Case-Based Reasoning (CBR), capable of being addressed alongside the design and implementation of advanced visualization tools.

This set of intelligent contributions translates into a total personalization of the reports, and has the ultimate purpose of facilitating the cognition and detection of relationships between patterns within the cases treated by advanced data visualization techniques, in addition to enabling the automation of decisions based on the observation and evaluation of the data associated with these cases. Therefore, it is of special interest to transfer this objective to Medicine and apply AI techniques such as CBR in systems composed of electromyography (EMG) sensors, in order to explain to the experts the characteristics of patient data and thus facilitate day-to-day decision making.

This Master's Final Project forms an important division within an ambitious research project, and the main objective defined for its implementation is based on the development of a prototype of an intelligent mechanism for generating customizable and interactive visual explanations with the ability to learn from interaction. Although the mechanism and methodology are established in a generic way, the proof of concept is made with data corresponding to the domain of Sports Medicine, specifically providing support to the medical professional in the process of interpreting characteristics of data associated with patients that correspond to electromyography analysis.

To do this, first of all, the CBR processes and hierarchies in the reuse of visual elements in reports are described. Next, the nature of the data visualization techniques is analysed in a generic way, in order to subsequently focus on a specific entity profile within the field of Sports Medicine. This approach is approached throughout the project as a practical study, and at all times the aim is to exploit the functionality of the tool belonging to this entity in order to then enhance its evolution from both a descriptive and predictive perspective. This analysis of medical data ultimately leads to a review of the legislative procedures for the processing of the data themselves in the health field.

**Keywords:** *Visualization, CBR, IA, Big Data, Sports Medicine, EMG, storytelling, interaction, explanations, analysis, knowledge, reuse, adaptation, learning, graphics, automation, legality.*

## **Agradecimientos**

Quisiera expresar mi más profundo reconocimiento y agradecimiento a todas aquellas personas que, gracias a su colaboración, han contribuido a la realización de este Trabajo Final de Máster.

En primer lugar, quisiera agradecer a Belén Díaz Agudo y Juan Antonio Recio García, los directores, la oportunidad que me han brindado de llevar a cabo este ambicioso proyecto, desarrollado dentro del grupo de investigación GAIA de la UCM (Group 921330) y del marco del proyecto TIN2017-87330-R, RAZONAMIENTO BASADO EN CASOS PARA LA EXPLICACION DE SISTEMAS INTELIGENTES, del MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD.

Gracias a Ignacio y todo el equipo de mDurance, por ser una fuente proactiva de colaboración en este proyecto y sobre todo por tener una propuesta de valor para el bien de la sociedad.

En general, gracias a todos los profesores de los cuales he tenido el honor de aprender en el último año, bien sea desde el punto de vista académico o bien como referentes profesionales. Es un privilegio pertenecer a la primera promoción del Máster Universitario de Internet de las Cosas, y personalmente considero casi un deber agradecer a quién ha puesto todo su empeño y esfuerzo en promover el aprendizaje en un área con semejante potencial.

Por último y muy especialmente, gracias a mi familia y a Beatriz, por comprender y animarme como nadie en esta breve pero intensa etapa que me ha hecho crecer como profesional y, con todo su apoyo, también como persona.

A todos, MUCHAS GRACIAS.



# Índice

Capítulo 1. Introducción .....	1
1.1    Objetivos del Trabajo Final de Máster .....	3
1.2    Estructura del Trabajo Final de Máster .....	3
Introduction .....	4
Objectives of the Final Master’s Project .....	6
Structure of the Final Master’s Project .....	6
Capítulo 2. Estado del arte .....	7
2.1    Tratamiento masivo de datos .....	7
2.1.1    Concepto de Big Data .....	7
2.1.2    ¿Quién hace uso de Big Data y quién no? .....	8
2.2    Sistemas Basados en Conocimiento .....	8
2.2.1    Fundamento del CBR .....	9
2.2.2    CBR como proceso y jerarquía .....	10
2.3    La visualización de datos como inteligencia .....	11
2.3.1    Percepción humana en DataViz .....	11
2.3.2    Principios de diseño visual .....	12
2.3.3    La teoría del color .....	14
2.3.4    Elementos gráficos en DataViz .....	18
2.3.5    Storytelling: la importancia del contexto .....	21
2.4    Ámbito de aplicación .....	22
2.4.1    Análisis y representación de datos en el ámbito sanitario .....	22
2.4.2    La electromiografía como ciencia .....	23
2.4.3    A quién se dirige .....	23
2.4.4    Responsables del análisis médico .....	24
2.4.5    Alcance tecnológico del análisis .....	24
2.5    Recapitulación del estado de la materia .....	25
Capítulo 3. Metodología .....	26
3.1    Alcance del desarrollo .....	26
3.2    Generación de plantillas de informes .....	26

3.3	Sistemas de automatización de datos .....	28
3.3.1	Ciclo CBR para la finalización de informes .....	28
3.3.2	Jerarquía CBR para la finalización de informes .....	29
3.3.2.1	Tarea de recuperación.....	30
3.3.2.2	Tarea de reutilización .....	31
3.3.2.3	Tarea de revisión .....	32
3.3.2.4	Tarea de retención .....	33
3.4	Caracterización de datos.....	35
3.4.1	Etiquetas sintácticas .....	36
3.4.1.1	Etiquetas estadísticas .....	36
3.4.1.2	Etiquetas probabilísticas .....	37
3.4.2	Etiquetas semánticas .....	39
3.5	Caracterización de gráficos .....	39
3.5.1	Gráficos de barras .....	41
3.5.2	Gráficos de líneas.....	42
3.5.3	Gráficos de caja.....	43
3.5.4	Gráficos de dispersión.....	44
3.5.5	Gráficos de burbujas .....	45
3.5.6	Gráficos de pastel.....	46
3.5.7	Diagrama de Venn .....	47
3.5.8	Mapa de árbol .....	48
3.5.9	Mapa geográfico .....	50
3.6	Recapitulación de la metodología genérica propuesta .....	51
Capítulo 4. Caso práctico de estudio y aplicación .....		52
4.1	mDurance .....	52
4.2	Exploración del contenido de la plataforma as-is .....	53
4.2.1	Exploración de los modelos de informes iniciales.....	53
4.2.2	Exploración de contenido de datos .....	54
4.2.3	Exploración de elementos visuales .....	55
4.3	Experimentación.....	58
4.3.1	Características del estudio.....	58
4.3.2	Características de la población.....	58
4.3.3	Instrumentación y tecnologías .....	59

4.3.3.1	Dispositivo IoT utilizado .....	59
4.3.3.2	Tecnologías empleadas.....	60
4.3.4	Movimientos .....	62
4.3.4.1	Pruebas de movimientos isotónicos.....	63
4.3.4.2	Pruebas de movimientos isométricos .....	64
4.4	Procesamiento de datos .....	65
4.4.1	Filtrado.....	67
4.4.2	Eliminación de ruido.....	68
4.5	Análisis de datos.....	68
4.6	Etiquetado de datos .....	71
4.6.1	Etiquetas sintácticas .....	72
4.6.2	Etiquetas semánticas .....	74
4.7	Optimización de la plataforma, estado to-be.....	76
4.7.1	Modelos de informes propuestos .....	77
4.7.2	Ejemplo general de utilización de los informes.....	77
4.8	Diseño de elementos visuales objeto de reutilización y adaptación.....	79
4.8.1	Estudio de la población de pacientes .....	80
4.8.2	Estudio del estado físico de un paciente .....	87
4.8.3	Comparación entre dos pacientes.....	93
4.8.4	Comparación entre un paciente y el resto de población .....	96
4.8.5	Ejemplo pormenorizado de utilización de los informes .....	100
Capítulo 5. Legalidad en el contexto .....		105
5.1	Anonimización de los datos .....	105
5.2	Punto de partida.....	105
5.3	Antecedentes reguladores: LOPD .....	106
5.4	Nueva regulación en vigor: GDPR.....	107
5.4.1	Consentimiento explícito .....	107
5.4.2	DPO .....	107
5.4.3	Aumento del nivel de información .....	108
5.4.4	Medidas organizativas y de seguridad.....	108
5.4.5	Evaluación del impacto.....	110
5.4.6	Comunicación de los datos personales .....	110

Capítulo 6. Conclusiones .....	111
6.1 Limitaciones .....	113
6.2 Próximos pasos.....	113
Conclusions .....	115
Limitations .....	117
Next steps .....	117
Bibliografía .....	118
Anexos .....	123
Anexo 1: Especificaciones del wearable Shimmer 3 EMG [71].....	123
Anexo 2: Paletas de color utilizadas .....	126

## Capítulo 1. Introducción

En ningún momento histórico como en el que nos encontramos se ha podido contemplar un auge de tal magnitud en la capacidad inteligente generada por agentes no humanos. Desde que científicos y matemáticos comenzaron a recoger las primeras inspiraciones allá por la década de 1950 y de forma totalmente ininterrumpida hasta hoy día, los sistemas de Inteligencia Artificial (IA) han evolucionado a un ritmo que nos atrevemos a decir no tiene comparación con ningún otro fenómeno tecnológico.

Ya durante los años 70 y 80 se hacían realidad múltiples desarrollos y simulaciones que abrían nuevas vías de investigación para estas técnicas, y fue algo antes de la entrada del nuevo milenio cuando las principales empresas se vieron metidas casi de forma inintencionada en una carrera tecnológica que avanzaba sin precedentes de cara a esta nueva industria. Ninguna compañía quería perder detalle en las ramas referentes a la teoría de la información, redes neuronales, computación, abstracción y creatividad [1].

No obstante, si bien es cierto que son muchos los paradigmas que se arrastran en esta ciencia a lo largo de estas tres últimas décadas, echando la vista unos pocos años atrás nos topamos con uno de los acontecimientos que ha resultado ser el detonante: el aumento masivo de las cantidades y variedades de los conjuntos de datos.

Uno de los causantes clave en esta *revolución de los datos* [2] es la forma en la que se ha hecho efectiva la incorporación de los mismos en el mundo físico. Hasta hace relativamente poco acostumbrábamos a recoger conjuntos de datos únicamente desde el entorno web, donde las virtudes y complejidades de los procesos en cierto modo son conocidos. Actualmente, la procedencia de los datos deriva de flujos de todo tipo y destaca cada vez más la constante aportación del entorno Internet of Things (IoT) gracias a sensores y wearables que, utilizados en conjunto con otras plataformas Cloud, dejan atrás la utilización de mecanismos tradicionales para cualquiera de las fases de procesado, almacenamiento, análisis e interpretación.

De forma inevitable, este conjunto de tecnologías lleva consigo la utilización de sistemas inteligentes capaces de aportar valor y veracidad a tal cantidad de datos mediante capacidades de percepción, razonamiento, decisión y actuación autónoma sobre dichos conjuntos de datos. Desafortunadamente, en todo sistema inteligente de estas características existe un hándicap que pone la confianza del usuario humano en contra del razonamiento artificial, y no es otra cosa que el comportamiento de los sistemas a modo de cajas negras [3], lo que se traduce en cierta incapacidad a la hora de explicar las acciones y decisiones tomadas por parte de estos sistemas, así como el hecho de poder justificar las recomendaciones y predicciones que autónomamente llevan a cabo.

Pues bien, es debido a este conjunto de incapacidades por las que el usuario humano requiere de una necesidad adicional que le lleve a entender las razones mediante las cuales un sistema de IA llega a una conclusión sobre el contexto. Para tratar de solventar esto, en primer lugar, es imprescindible asegurar la calidad del dato [4] como pilar fundamental del procedimiento. De esta forma es posible hacer un uso eficiente del Razonamiento Basado en Casos (CBR) que permite incorporar explicaciones en última instancia sobre el contexto.

Por tanto, gracias al paradigma CBR se puede dotar de sentido tanto la inmensidad de datos heterogéneos recogidos por sensores IoT como los generados en la aplicación de

técnicas de IA en distintos entornos. Así pues, una forma de optimizar el razonamiento en estas áreas – IoT e IA – está enfocada a la generación de informes que permitan explicar la inmensa cantidad de información generada según un contexto determinado y de la forma más visual posible hacia el usuario humano.

Cabe decir que es precisamente en este punto donde el campo de Data Visualization (DataViz) juega un papel crucial, al estar dedicado a la investigación de representaciones gráficas encargadas de facilitar al usuario humano el reconocimiento y detección de relaciones entre patrones entre conjuntos de datos. Además, esta área tiende a ir más allá pues se caracteriza por su habilidad para revelar y contar historias con los datos. El llamado *storytelling* [5] es un medio de comunicación cada vez más utilizado para la obtención de ideas sobre conjuntos formados por tres componentes: datos, elementos visuales y parte narrativa. Muy adecuado, por tanto, para la generación de informes.

Con todo lo anterior, se puede afirmar que el conjunto de sistemas IoT e IA complementados con las técnicas CBR y DataViz permiten obtener conocimiento en un contexto y hacer predicciones sobre ocurrencias futuras. Esto hace que este conjunto se convierta casi de forma automática en un caso de éxito independientemente del campo de aplicación en el que se ubique, pudiéndose trasladar a cualquier mercado [6] donde los beneficios se traducen en incrementos de productividad multiplataforma y multicanal gracias, por un lado, a la automatización de procesos y el ahorro en consumos; y por otro, al incremento del grado de personalización de servicios, que aumenta proporcionalmente al número de interacciones del usuario con el sistema.

No obstante, es muy importante contemplar esta ciencia más allá del punto de vista del área de negocios, pues a nivel de aplicación puede trabajar transversalmente con otras ramas de la ciencia y ser determinante en necesidades sociales de primer nivel como bien es el ámbito sanitario [7]. Aquí, los profesionales y responsables encargados de la prevención de riesgos médicos pueden beneficiarse de la alta eficiencia en la recogida de datos mediante sistemas IoT así como de la aplicación de técnicas inteligentes para proporcionar, entre otros, informes visuales que ayuden a crear un mayor nivel de transparencia acorde a los principios éticos en la toma de decisiones en el campo de la Medicina.

A esto se le ha de sumar la amplia cantidad de dominios donde el uso de CBR ha supuesto una tremenda capacidad de resolver problemas complejos recordando experiencias previas, llegando a abordar múltiples sistemas de reconocimiento de patrones y de soporte en la toma de decisiones.

De este modo, desde el punto de vista del experto Médico, el conjunto de sistemas inteligentes con CBR y herramientas avanzadas de visualización no solamente crea fiabilidad en la obtención del conocimiento sobre estos datos, sino que también sirve de precedente para aplicar unos algoritmos de análisis u otros y de forma simultánea aprovechar la interacción del profesional para dar retroalimentación al sistema y así crear predicciones sobre lo que podrá ocurrir o no en el futuro. Este caso es particularmente útil en el dominio de aplicación de la electromiografía, donde los datos recogidos de múltiples sensores a priori carecen de sentido y suponen en la mayoría de ocasiones una fuente de preocupación entre los expertos en la materia.

Claramente estamos ante un área con un potencial tremendo y que la sociedad en general es capaz de aprovechar muy positivamente. La explicación de datos mediante informes visuales y textuales en el campo de la Medicina ayuda tanto al experto como al usuario a comprender los detalles de los patrones establecidos en pruebas a distintos

niveles, así como a interpretar de manera automática, fiable y precisa los enormes conjuntos de datos en base a un contexto.

## **1.1 Objetivos del Trabajo Final de Máster**

El objetivo de este Trabajo Final de Máster se basa en la prueba de concepto de un prototipo de herramienta inteligente capaz de generar explicaciones visuales e interactivas dentro de informes, particularizando a posteriori en la interpretación por parte de los profesionales médicos sobre las características de los datos asociados a pacientes. Además, también se persigue dotar de capacidad de aprendizaje a dicha herramienta mediante la interacción del propio experto de modo que el sistema sea capaz de aumentar su grado de personalización.

Se propone desarrollar un prototipo de herramienta inteligente de generación de explicaciones visuales personalizables e interactivas con capacidad de aprendizaje a partir de la interacción. Aunque la herramienta y la metodología se establecen de forma genérica, la prueba de concepto se realiza con datos correspondientes al dominio de la Medicina del Deporte.

Para ello, el Trabajo Final de Máster se puede desglosar en una serie de objetivos específicos, que se pueden resumir en:

- Presentación de los fundamentos que conforman el estado del arte de la materia a tratar en el transcurso de todo el proyecto.
- Análisis del marco teórico que sigue la metodología CBR en la generación de explicaciones interactivas dentro de informes dependiendo de las características genéricas de los modelos de datos.
- Estudio del alcance del proyecto y metodología seguida en la recomendación automática de elementos de visualización sobre una plantilla de informes y un modelo de datos caracterizado en cualquier ámbito de aplicación.
- Diseño de técnicas inteligentes para la personalización de informes dentro del campo de la Medicina del Deporte, capacitando el aprendizaje a partir de la observación y evaluación de datos asociados a pacientes.
- Descripción de los procedimientos legales tenidos en cuenta a la hora de tratar y representar datos de carácter personal y especialmente protegidos.

## **1.2 Estructura del Trabajo Final de Máster**

Este documento se estructura en un total de 6 capítulos. El primero de ellos consiste en el presente capítulo introductorio del Trabajo Final de Máster, y el segundo abarca la situación de la materia a tratar, en líneas generales. Seguidamente, mediante el tercer capítulo se tratan los conceptos a considerar tanto en la metodología CBR del proyecto global como la parte concerniente y aplicable en este Trabajo Final de Máster.

Así pues, el cuarto capítulo entra en materia práctica mediante el desarrollo de técnicas para la generación de explicaciones visuales particularizadas en el dominio de la Medicina del Deporte, concretamente con datos procedentes de análisis EMG, con tal de conseguir un alto nivel de personalización en plantillas de informes enfocados a casos de pacientes.

Para finalizar, se da paso al estado legislativo que respecta la utilización y el tratamiento de datos dentro de este campo de la Medicina del Deporte; y por último, en el sexto capítulo se exponen las conclusiones obtenidas a raíz de este proyecto.

## Introduction

At no time in history has it been possible to contemplate a boom of such magnitude in the intelligent capacity generated by non-human agents. Since scientists and mathematicians began to collect the first inspirations back in the 1950s and completely uninterruptedly until today, Artificial Intelligence (AI) systems have evolved at a pace that we dare to say has no comparison with any other technological phenomenon.

Already during the 70's and 80's, multiple developments and simulations became reality that opened new ways of researching for these techniques, and it was some time before the entry of the new millennium when the main companies were almost unintentionally involved in a technological race that was advancing without precedent in the face of this new industry. No company wanted to lose any single detail in the branches of information theory, neural networks, computing, abstraction and creativity [1].

However, while it is true that there are many paradigms that drag this science through these last three decades, looking back a few years we came across one of the events that has turned out to be the trigger: the massive increase in the quantities and varieties of data sets.

One of the key causes of this data revolution [2] is the way in which the incorporation of data into the physical world has become effective. Until relatively recently we used to collect data sets only from the web environment, where the virtues and complexities of the processes are somehow known. On the other hand, currently, the origin of the data derives from all kinds of flows and the constant contribution of the Internet of Things (IoT) environment is increasingly highlighted thanks to sensors and wearables that, used in conjunction with other Cloud platforms, leave behind the use of traditional mechanisms for any of the processing, storage, analysis and interpretation phases.

Unavoidably, this set of technologies entails the use of intelligent systems capable of adding value and truthfulness to such a quantity of data by means of capacities of perception, reasoning, decision and autonomous action on these sets of data. Unfortunately, in any intelligent system of these characteristics there is a handicap that turns the trust of the human user against artificial reasoning, and it is nothing other than the behavior of the systems in the form of black-boxes [3], which translates into a certain inability to explain the actions and decisions taken by these systems, as well as the fact of being able to justify the recommendations and predictions that they autonomously carry out.

Well, it is because of this set of inabilities that the human user requires an additional need to understand the reasons by which an AI system reaches a conclusion about the context. In order to solve this, first of all, it is essential to ensure data quality [4] as a fundamental pillar of the procedure. In this way, it is possible to make efficient use of Case Based Reasoning (CBR), which allows the incorporation of explanations over the context.

Therefore, thanks to the CBR paradigm, both the immensity of heterogeneous data collected by IoT sensors and those generated in the application of AI techniques in different environments can be given meaning. Thus, one way to optimize reasoning in these areas - IoT and AI - is focused on the generation of reports that can explain the immense amount of information generated according to a given context and in the most visual way possible to the final human user.

It is precisely at this point where the Data Visualization (DataViz) field plays a crucial role, as it is dedicated to the investigation of graphic representations in charge of facilitating the human user the recognition and detection of relationships between patterns between data sets. In addition, this area tends to go further as it is characterized by its ability to reveal and tell stories with the data. Storytelling [5] is a mean of communication increasingly used to obtain ideas about data sets composed by three components: data, visual elements and narrative part. Very suitable, therefore, for the generation of reports.

With all the above, it is said that the conglomeration of IoT and AI systems complemented with CBR and DataViz techniques allow to obtain knowledge in a context and make predictions about future occurrences. This means that this kind of conglomeration becomes almost automatically a success case regardless of the field of application in which it is located, being able to be transferred to any market [6] where the benefits are translated into cross-platform and cross-channel productivity increases thanks, on the one hand, to the automation of processes and savings in consumption; and on the other hand, to the increase in the personalization services' degree, which extends proportionally to the number of interactions between the user and the system.

Nevertheless, it is very important to contemplate this science beyond the point of view of the business area, since at the application level it can work transversally with other branches of science and be a determining factor in some first-level social needs such as the health field [7]. Here, professionals and those responsible for the prevention of medical risks can benefit from the high efficiency in the collection of data through IoT systems as well as the application of intelligent techniques to provide, among others, visual reports that help to create a higher level of transparency in line with ethical principles in decision-making in the field of medicine.

It also must be added the large number of application domains where the use of CBR has meant a tremendous ability to solve complex problems by remembering previous experiences, going so far as to deal with multiple systems of pattern recognition and support in decision making.

Thus, from the point of view of the medical expert, the set of intelligent systems with CBR and advanced visualization tools not only creates reliability in obtaining knowledge about these data, but is also a good precedent to apply some analysis algorithms and simultaneously take advantage of the professionals' interaction to give feedback to the system and thus create predictions about what may or may not happen in the future. This case is particularly useful in the field of application of electromyography, where all the collected data from multiple sensors a priori are meaningless and in most cases a source of concern among experts in the field.

Clearly, we are dealing with an area with tremendous potential that society in general is capable of exploiting very positively. The explanation of data through visual and textual reports in the field of Medicine helps both the expert and the user to understand the details of the patterns established in tests at different levels, as well as to interpret automatically, reliably and accurately the huge data sets based on a context.

## **Objectives of the Final Master's Project**

The objective of this Master's Final Project is based on the proof of concept of an intelligent tool prototype capable of generating visual and interactive explanations within reports, subsequently specializing in the interpretation by medical professionals of the characteristics of the data associated with patients. In addition, it also aims to provide learning capacity to this tool through the interaction of the expert himself so that the system is able to increase its degree of customization.

It is proposed to develop a prototype of an intelligent tool for generating customizable and interactive visual explanations with the capacity to learn from interaction. Although both tool and its methodology are established in a generic way, the proof of concept is carried out with data corresponding to the domain of Sports Medicine.

To this end, the Master's Final Project can be broken down into a series of specific objectives, which can be summarized as follows:

- Presentation of the fundamentals that make up the matter of the subject to be dealt through the project.
- Analysis of the theoretical framework that follows the CBR methodology in the generation of interactive explanations within reports depending on the generic characteristics of the data models.
- Study of the scope of the project and methodology followed in the automatic recommendation of visualization elements on a report template and a data model characterized in any field of application.
- Design of intelligent techniques for the personalization of reports within the field of Sports Medicine, enabling learning from the observation and evaluation of data associated with patients.
- Description of the legal procedures taken into account when processing and representing personal and specially protected data.

## **Structure of the Final Master's Project**

This document is structured in a total of 6 chapters. The first of them consists of the present introductory chapter of the Final Master's Project, and the second covers the situation of the subject to be dealt with, in general terms. Next, the third chapter deals with the concepts to be considered both in the CBR methodology of the global project and the part concerning and applicable to this Final Master's Project.

Thus, the fourth chapter enters into practical matter through the development of techniques for the generation of particularized visual explanations in the field of Sports Medicine, specifically with data from EMG analysis, in order to achieve a high level of customization in templates of reports focused on patient cases.

Finally, the legislative status is explained regarding the use and processing of data within this field of Sports Medicine; and finally, in the sixth chapter, the conclusions obtained as a result of this project are presented.

## Capítulo 2. Estado del arte

Independientemente de cuál sea el nivel de abstracción de la disciplina Data Visualization, es completamente necesario analizar el estado del arte de esta rama y todas las influencias dentro del ámbito de las tecnologías de la información y la comunicación.

Es por ello que en el presente capítulo se realiza un recorrido teórico partiendo del concepto de Big Data en la sección 2.1 y se da paso a los Sistemas Basados en Conocimiento en el apartado 2.2, en concreto del CBR como metodología perteneciente a la rama de Inteligencia Artificial. Además, en el apartado 2.3 se realiza un estudio de DataViz como ciencia y forma de conocimiento y; por último, en la sección 2.4, se abre paso al contexto que va a ser tratado en los siguientes capítulos, la electromiografía como Medicina del Deporte.

### 2.1 Tratamiento masivo de datos

Hoy en día el concepto Big Data se encuentra presente en cada movimiento que la sociedad hace frente dentro del ámbito de las Tecnologías de la Información. Según la materia que resulte de interés para el portavoz, el enfoque puede llegar a variar, llegando a tener incluso definiciones contradictorias.

#### 2.1.1 Concepto de Big Data

La descripción que se arrastra desde hace unos años [8] habla de “*todo conjunto datos (estructurados, no estructurados y semiestructurados) con información que no puede ser procesada o analizada mediante procesos o herramientas tradicionales*”.

En base a esta definición podrían surgir una fuente de discrepancias en función de tres aspectos, fundamentalmente:

- En primer lugar, el *volumen* de la información, ya que no se podría especificar siquiera una magnitud concreta en la que se pudiera aclarar el término “Big”.
- Por otro lado, la *variedad* en la que se disponen los datos, bien sea por la tipología en la que se presentan o por las fuentes de datos cada vez sobre tecnologías más avanzadas.
- Seguidamente, la *velocidad* y movimiento de procesado de datos, debido a la heterogeneidad de herramientas emergentes con distinta capacidad para abarcar servicios.

Es por ello que la definición nombrada anteriormente puede crear cierta ambigüedad sobre cualquier entidad o institución que pretenda hacer uso del Big Data. A la hora de examinar un posible diseño, desconocería la cantidad mínima de datos y carecería de una especificación sobre fuentes de información, límites de tasas de velocidad en cualquiera de las fases del tratamiento de los datos.

### 2.1.2 ¿Quién hace uso de Big Data y quién no?

No obstante, han sido precisamente las tres fuentes de discrepancias nombradas las que han ido dejando atrás este término inicial y han condicionado su evolución, junto al constante crecimiento de herramientas de software de código abierto utilizadas en organizaciones públicas y privadas pertenecientes a sectores altamente dispares. Esto ha provocado paulatinamente que hoy en día el término Big Data sea un reto mayor, y posea la madurez suficiente como para ser considerado más bien como un conjunto de tecnologías, algoritmos y sistemas empleados para la recolecta de datos a una escala y variedad no alcanzada hasta ahora, con el valor añadido de posibilitar extracciones de información de valor mediante sistemas analíticos avanzados.

Este nuevo enfoque tecnológico adquirido en los últimos años quiere decir que hoy en día no se considera únicamente Big Data los cientos de Zettabytes de datos que procesan y almacenan información en formato digital, sino que se pretende ir más allá y focalizar en el pequeño porcentaje de datos que son útiles entre estos inmensos conjuntos y que permiten realmente proporcionar conocimiento [9].

De modo que:

- Por un lado, habría que tener en cuenta la *veracidad* de los datos que se disponen, pues en múltiples ocasiones se tiene un concepto equivocado de lo que realmente se tiene una vez se realiza cualquier display.
- Otra consideración surge en si los datos recogidos y representados realmente tienen o no *valor*, ya que de aquí pende la utilidad final de los mismos y el conocimiento que puede aportar realmente.

Ésta sería, por tanto, una de las justificaciones sobre las que se apoyan un elevado número de entidades e instituciones ya que, a pesar de no tratar las mayores cantidades de datos de forma absoluta, sí hacen uso realmente del Big Data debido a las herramientas y tecnologías que implican la transmisión de información fiable y con capacidad de aportar conocimiento a un grupo de usuarios finales. De cualquier forma, siempre surge la verdadera necesidad de comprender los datos que se han adquirido y, por consiguiente, de sistemas que generen explicaciones y supongan una fuente de valor y veracidad.

## 2.2 Sistemas Basados en Conocimiento

En relación a lo anterior cabe decir que el uso de sistemas Big Data no es factible sin un propósito final capaz de aportar conocimiento. Hoy día es muy común disponer de un entorno donde el IoT, Big Data, y las técnicas de IA formen un sólido equipo [10] y se complementen entre sí para generar valor descriptivo y predictivo a los conjuntos de datos:

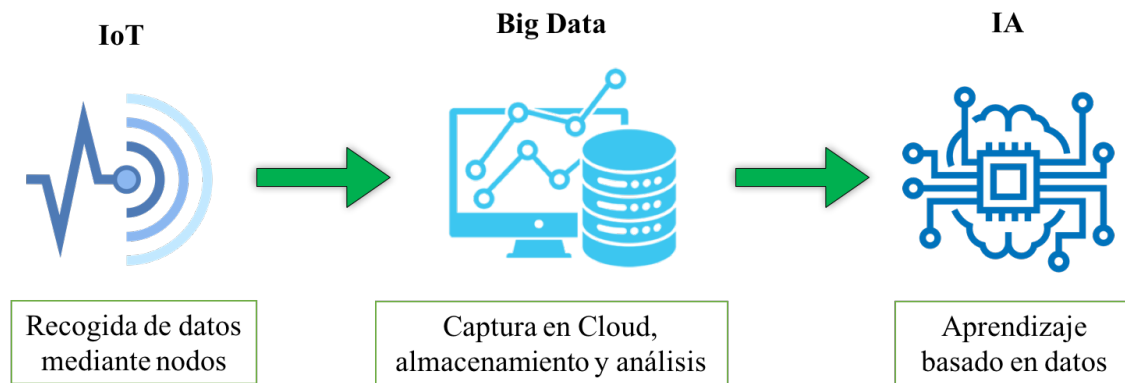


Figura 1. Sistema data-driven para la obtención y regeneración de conocimiento.

Este campo de sinergias es totalmente imprescindible ya que, en primera instancia, los datos recogidos por los microcontroladores IoT no tienen sentido si no disponen de un acceso Cloud donde se realice un análisis inmediato de los mismos y sean agrupados con otras fuentes de datos. Las tecnologías Big Data, por su parte, son las encargadas de crear un entorno IoT en conjunto con otros muchos flujos para realizar segmentaciones eficaces y así poder aplicar técnicas de IA que lleven a cabo una tarea considerada como esencial en la actualidad: la interpretación de datos para la **resolución de problemas**.

Este término de resolución de problemas se toma en este proyecto en el sentido amplio dentro de la Ingeniería del Conocimiento [11], una rama de la IA concerniente a la extracción y representación de conocimientos que normalmente requieren un alto nivel de experiencia humana. Esta rama se encarga de diseñar y hacer funcionar los Sistemas Basados en Conocimiento (SBC), capaces de mejorar las prestaciones de otros sistemas mediante la incorporación de conocimiento y sin necesidad de manipular la estructura global del sistema en cuestión. Esto es realmente valioso ya que permite obtener varias soluciones y razonar cómo se llega a dicha solución.

Así pues, un SBC procesa la información recogida y la incorpora a los recursos de razonamiento interpretables por un humano o por una máquina. Para ello, es muy común utilizar una memoria permanente en la cual se almacenen de manera explícita las soluciones dentro del dominio de aplicación en el que se trabaje. Este método en realidad es una de las formas más eficaces de razonamiento y aprendizaje dentro de los SBC, y no es otro que el paradigma Case-Based Reasoning (CBR) [12].

### 2.2.1 Fundamento del CBR

El CBR es una técnica de IA cuyo funcionamiento es equivalente a un sistema de razonamiento por analogía, ya que reutiliza casos anteriores para facilitar la adquisición de conocimiento. Esto tiene como justificación el hecho de que el ser humano tiende a resolver problemas en base a experiencias pasadas, y no a partir de un conocimiento explícitamente detallado, por lo que se puede afirmar que el CBR comprende una poderosa fuente de resolución de problemas para el humano.

En la terminología CBR, un caso supone una situación de problema y, por consiguiente:

- Una situación experimentada previamente es denominada como un caso pasado, previo, almacenado o retenido.
- El conjunto de problemas resueltos es la Base de Casos.

- Un nuevo caso es toda aquella situación del nuevo problema a resolver.

### 2.2.2 CBR como proceso y jerarquía

Por la suma de lo anterior, el CBR es considerado como un proceso cíclico, ya que se involucra en la recurrencia de los problemas y en el aprendizaje a partir de los mismos única y exclusivamente cada vez que el sistema lo necesita (metodología Lazy Learning). Para que esto se haga efectivo el ciclo CBR se describe según cuatro procesos [13]:

1. *Recuperación* de casos similares dentro de la Base de Casos.
2. *Reutilización* de la información y conocimiento en el caso capaz de resolver el problema.
3. *Revisión* de la solución propuesta.
4. *Retención* de las partes dentro de la resolución del problema que pueden ser útiles en experiencias posteriores.

Para poder interpretar este proceso secuencial es necesario orientar el sistema CBR de modo que la resolución de tareas siga un orden jerárquico donde cada subproceso es interpretado por el sistema como un caso o problema. Esta jerarquía describe un sistema CBR desde tres perspectivas:

- Como punto de partida, las **tareas** que comprenden los objetivos del sistema.
- Cada una de esas tareas se compone de uno o más **métodos**, y para que un método sea capaz de cumplir una tarea, necesita:
- **Conocimiento** sobre el dominio de aplicación, así como del problema en cuestión y su contexto.

De este modo, la tarea de nivel superior es la resolución de problemas y el aprendizaje a partir de la experiencia, que se implementa por el método CBR para así descomponerse en cuatro tareas de segundo nivel que se corresponden con las cuatro fases del proceso CBR: recuperación, reutilización, revisión y retención. Estas cuatro tareas son las que consiguen llevar a cabo la tarea de orden superior, y del mismo modo se segmentan en otras tareas de tercer nivel y así consecutivamente.

Cabe decir que, a partir de las tareas que suponen cada una de las fases del ciclo CBR no existen unos métodos predefinidos para cada dominio de aplicación, sino que el hecho de establecer esta jerarquía supone precisamente el reto de los sistemas CBR.

## 2.3 La visualización de datos como inteligencia

Como ya se ha comentado, uno de los puntos más importantes dentro de la revolución Big Data y los Sistemas Basados en Conocimiento consiste en la disposición de los conjuntos de datos a la hora de transmitir y comunicar la información, para así poder aportar conocimiento tanto al ser humano como a otros sistemas sobre tendencias descriptivas, iniciativas analíticas y modelos predictivos [14].

Bien es cierto que el camino del dato es rigurosamente complejo y hemos visto que depende de múltiples entornos. No obstante, en última instancia, la obtención de un valor analítico efectivo tiene como responsable final la ciencia DataViz, pues es un hecho que ofrece una forma de generar explicaciones y ampliar el conocimiento mucho más fácil que mediante cualquier otro modelo estadístico de computación convencional.

Es por ello que en los siguientes subapartados se analiza en profundidad la importancia de la percepción visual humana, la influencia de los elementos visuales en la transmisión de información, y por último algunas de los métodos generalmente más efectivos en las explicaciones gráficas.

### 2.3.1 Percepción humana en DataViz

La generación de explicaciones visuales tiene como foco de origen el propio procesamiento de la atención humana, que distingue entre una fase pre-consciente [15] donde se observan características como el *color*, la *forma*, la *textura*, el *movimiento* y la *posición* de los objetos en un intervalo temporal de menos de 200ms; con otra fase consciente en la cual el ser humano focaliza la atención de manera objetiva en una cantidad de tiempo mucho mayor.

Es por esto que la ciencia de DataViz se hace efectiva cuando distribuye eficazmente la percepción y la cognición humana [16]. Si analizamos cómo se comportan estas dos partes en nuestro cerebro, se obtienen las siguientes conclusiones:

- La *percepción visual* es manejada por la corteza de cerebro posicionada en la parte trasera, que es extremadamente rápida y eficiente. Es por eso que la vista actúa de forma inmediata y con poco esfuerzo.
- La *percepción cognitiva*, por otro lado, es manejada por la parte delantera del cerebro, que es mucho más lenta y menos eficiente. Es por ello que dotar sentido de forma aislada a un conjunto de datos requiere ocupar toda esta parte de cognición, y una carga de trabajo importante.

La visualización de datos consigue un balance entre estas dos partes, que se traduce en una visualización e interpretación instantánea y eficiente (sin mucha carga de trabajo) de los conjuntos de datos en nuestro cerebro. Se puede decir, por tanto, que la ciencia de DataViz permite obtener todas las ventajas de las habilidades cerebrales que engloban la percepción visual y cognitiva.

En la siguiente tabla se muestran ejemplos de atributos comúnmente utilizados a la hora de representar información, junto a sus capacidades de percepción humana. Dichas capacidades se miden por niveles en función del grado de asimilación en nuestro cerebro:


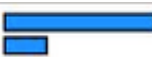












Display	Atributo	Comparación y Orden	Valores útiles	Cantidad	Número	Categorías	Relaciones
	Posición	Sí	Infinitos	Muy Bueno	Bueno	Bueno	Bueno
1, 2, 3; A, B, C	Etiquetas de texto	Sí	Infinitos	Muy Bueno	Bueno	Bueno	Bueno
	Longitud	Sí	Muchos	Muy Bueno	Bueno	N/A	N/A
	Tamaño	Sí	Muchos	Bueno	Bueno	N/A	N/A
	Orientación	Sí	Medio	N/A	Bueno	N/A	N/A
	Densidad	Sí	Pocos	Bueno	Bueno	N/A	N/A
	Grosor	Sí	Pocos	Bueno	Bueno	N/A	N/A
	Saturación, luminancia	Sí	Pocos	Bueno	Bueno	N/A	N/A
	Color	No	Muy pocos	N/A	N/A	Bueno	N/A
	Forma	No	Medio	N/A	N/A	Bueno	N/A
	Textura	No	Medio	N/A	N/A	Bueno	N/A
	Conexiones y cajas	No	Infinitos	N/A	N/A	Bueno	Bueno
	Modelo de línea	No	Pocos	N/A	N/A	N/A	Bueno
	Fin de línea	No	Pocos	N/A	N/A	N/A	Bueno
	Grosor de línea	Sí	Pocos	N/A	Bueno	N/A	N/A

Tabla 1. Atributos de codificación visual según la percepción humana y las capacidades que los cualifican [17].

Así pues, recapitulando para una capacidad de elevado interés como es la *percepción cuantitativa* que proporcionan los atributos visuales de la tabla anterior, contemplamos que (aparte de los números) atributos como la longitud y la posición son los que mejor precisión proporcionan, seguidos de la densidad, tamaño, saturación y brillo. No obstante, otros como la orientación, forma, cierre o conexiones no ofrecen percepción cuantitativa.

Por último, cabe decir que este tipo de codificación visual es efectiva exclusivamente si se considera de forma conjunta, ya que a la hora de generar elementos gráficos se agrupan gran parte de los atributos que aparecen en la tabla.

### 2.3.2 Principios de diseño visual

Una vez vistos los distintos tipos de codificación se puede contemplar la relación entre los patrones desde el punto de vista del diseño. Una teoría muy consistente acerca de la

percepción humana y la comunicación entre objetos se basa en los llamados principios de Gestalt [17]. Se resumen a través de la siguiente tabla:





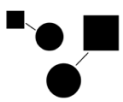
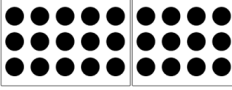
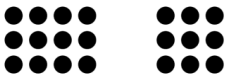
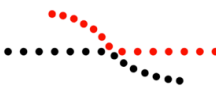
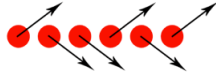


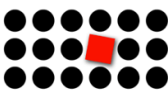

Ley	Explicación	Ejemplo
Simplicidad	El ser humano prefiere objetos simples, claros y ordenados, pues son concebidos más rápidamente y sin necesidad de reorganización visual.	
Cierre	Si se proporciona la cantidad exacta de información, el ojo humano es capaz de completar el resto por sí mismo y tratar a los distintos objetos o formas como un conjunto.	
Simetría y orden	La simetría precede a la proximidad, y eso se demuestra en formas que el ojo humano tiende a asimilar de forma rápida.	
Figura y fondo	Los objetos de forma convexa y de tamaño más pequeño tienden a ser considerados como figura, mientras que los de forma cóncava y de mayor dimensión se consideran como fondo.	
Conexiones	Los objetos conectados entre sí tienden a concebirse visualmente como un conjunto, aunque existan otros pares de objetos con su misma forma.	
Regiones comunes	Los objetos encerrados conjuntamente tienden a tratarse como un grupo a pesar de que se tengan objetos de su misma forma fuera de esa separación.	
Proximidad	Los objetos separados conjuntamente tienden a tratarse como un grupo a pesar de que se tengan objetos de su misma forma fuera de esa separación espacial.	
Continuidad	Los elementos con una disposición en línea recta o curva son concebidos con relación entre sí por la asimilación de los puntos origen-final.	
Sincronía	Los objetos orientados de la misma forma se conciben con relación entre sí por “dirigirse” a un mismo destino.	
Paralelismo	Los elementos que son paralelos entre sí se consideran que tienen relación o que se pueden tratar como un conjunto aparte del resto.	
Similitud	Los objetos que comparten características similares como forma, color, tamaño y textura son considerados por guardar una relación entre sí.	
Focalización	Los objetos con un punto particular de interés tienden a llamar la atención visual, aunque existan otros muchos de distintas características.	
Experiencias anteriores	Existen ciertos elementos visuales que se conciben de acuerdo a la experiencia con esas características, como por ejemplo asociación de colores.	

Tabla 2. Leyes de Gestalt sobre el diseño gráfico y la percepción humana.

Lo anterior define cómo el ser humano concibe los elementos visuales y de la forma que tiende a agruparlos o buscar relaciones entre ellos a la hora de ser representados. Esto es realmente útil a la hora de resaltar un conjunto de datos del resto y también a la hora de dotar de sentido a lo que a priori parece incoherente [18].

### 2.3.3 La teoría del color

El sentido que se le puede dotar al color está abierto a subjetividad en todo lo referente al diseño de elementos de visualización de datos. La posible reacción de un sujeto ante determinados conjuntos de colores puede variar en función de múltiples aspectos, partiendo de la edad y llegando incluso a factores éticos y culturales de dicho observador.

Desde el punto de vista de DataViz, la teoría del color [19] engloba el modelo HSL que es caracterizado por tres componentes claves: el propio *tono de color*, la *saturación* y la *luminosidad*. Estos se muestran a través de la figura 2 [20]:

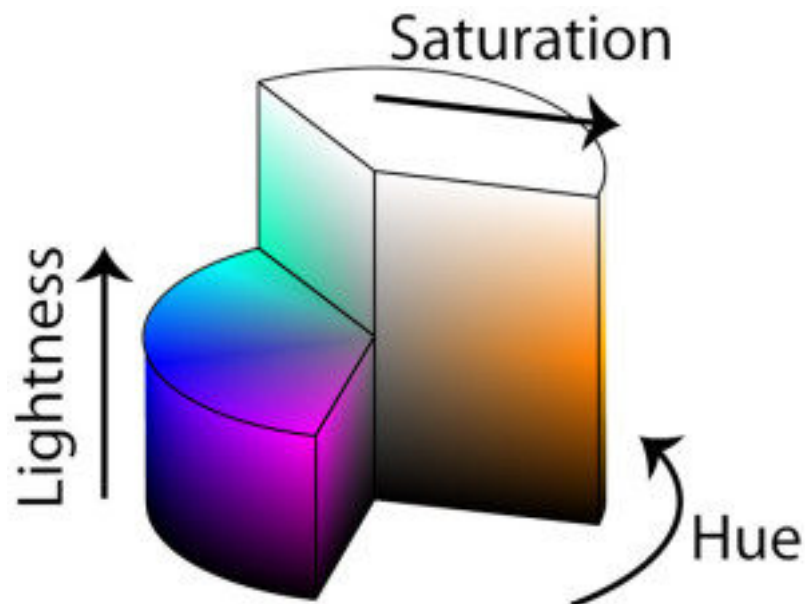


Figura 2. Componentes de color y sus respectivas escalas en el modelo HSL.

- El propio *tono de color* que se puede encontrar en la paleta de colores convencional, y que en este sistema se puede representar mediante una escala de 0° a 360° en forma de rueda.
- La *saturación*, que marca la intensidad del color. La escala HSL se encarga de medir la diferencia del color con el gris neutro, que tiene 0% de saturación. Los colores con un alto nivel de saturación parecen más brillantes y vivos.
- La *luminancia* describe el rango de variación de un tono de color desde la oscuridad hasta la claridad del mismo. El criterio para ello es la cantidad de color negro añadido.

El primer paso en la decisión de colores en una herramienta DataViz determinada es conocer los datos que se quieren representar mediante el color. Hay que ser coherente para poder determinar si realmente los atributos de datos son representables mediante

color. Tal y como se ha contemplado en la codificación humana en la percepción visual (ver tabla 1), el tono de color es muy apropiado para representar categorías, mientras que la saturación y la luminancia se asocian generalmente a valores cuantitativos.

Una vez se decide que un determinado atributo se va a representar mediante color, el siguiente paso es entender la escala de los datos correspondientes a dicho atributo para poder elegir los tonos de color. Generalmente se puede discernir entre tres tipos de series de datos en función de las cuales se puede hacer una estimación de la cantidad de colores a utilizar:

- **Datos secuenciales:** Cuando se parten de valores bajos que se conoce que tienden a ir hacia valores más altos. Este tipo de datos requieren un solo tono de color y la variación se puede llevar a cabo con el uso de la saturación y luminancia. No obstante, cualquier cambio sutil en estos dos componentes puede resultar difícil de apreciar, por lo que si se tiene un amplio rango de datos reducido se pueden utilizar dos tonos de color distintos.
- **Datos divergentes:** Cuando existen puntos determinantes en el principio y fin de la escala, así como otros puntos importantes justo en mitad de la distribución. Es por ello que este tipo de series de datos requieren dos tonos de color, descendiendo en saturación o luminancia hacia a un tono neutral de gris, blanco o negro.
- **Datos cualitativos:** Son aquellos que no tienen un orden de magnitud claro, y por tanto requieren tantos tonos de color como número de valores. Ante este tipo de serie de datos es importante tener en cuenta que la percepción visual humana comentada anteriormente tiene limitaciones al contemplar un elevado número de colores. Está demostrado que el uso 7 u 8 colores distintos ya requiere de percepción cognitiva además de visual, mientras que el uso de más de 12 provoca problemas de diferenciación en la capacidad humana.

Además, la elección de los propios tonos de color dentro de la paleta determina considerablemente la eficacia de la visualización. Esto se basa en la correcta combinación de los colores, pues es cierto que existen ciertos tonos que funcionan juntos mejor que otros. Hoy en día se pueden encontrar herramientas [21] que mediante una rueda de colores computan estas combinaciones en base a la *regla de color* que convenga para cada caso. Estas reglas de colores se establecen en base a una serie de criterios que se definen a continuación:

- La *escala monocromática* se encarga de jugar con las sombras dentro de un mismo tono de color, lo cual hace que sea ideal para las series de datos secuenciales.

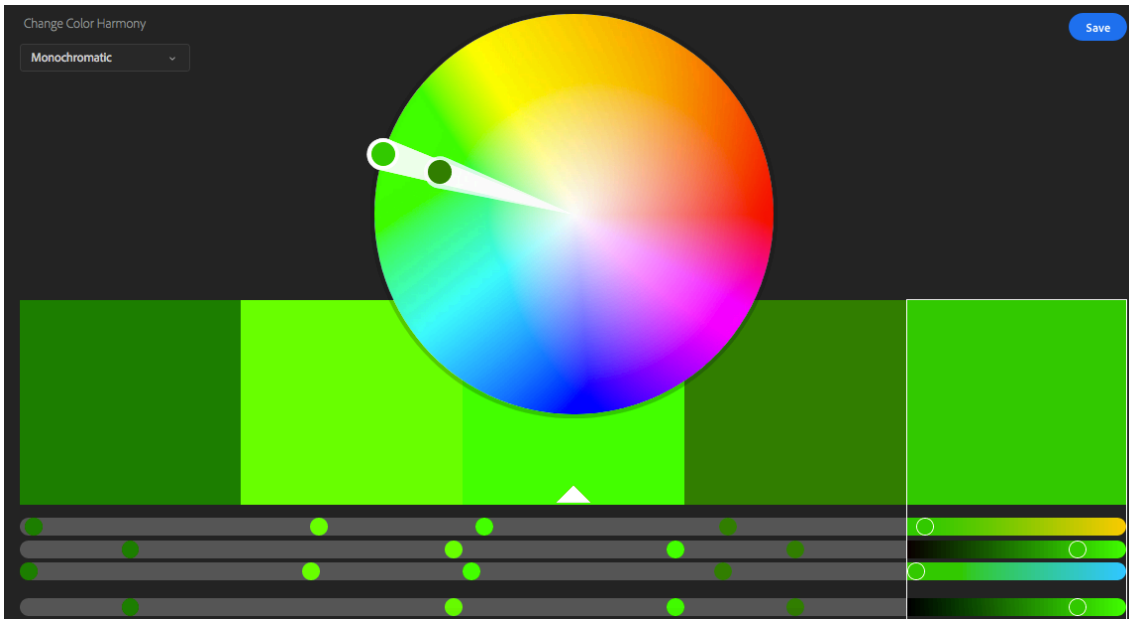


Figura 3. Ejemplo de aplicación de regla monocromática en el modelo HSL.

- La *escala de colores análogos* se encarga de establecer un alineamiento de colores entre sí dentro de la rueda de color, por lo que son una alternativa muy correcta desde el punto de vista estético en la representación de datos secuencial.



Figura 4. Ejemplo de aplicación de regla de analogía en el modelo HSL.

- La *escala de colores complementarios* son los que se sitúan en lados opuestos de la rueda. Éstos tienen como fortaleza el área de representación que se componga de dos tonos de color, aunque es cierto que si se emparejan con colores neutros (blanco o gris) suponen una combinación perfecta para la representación de datos divergentes.

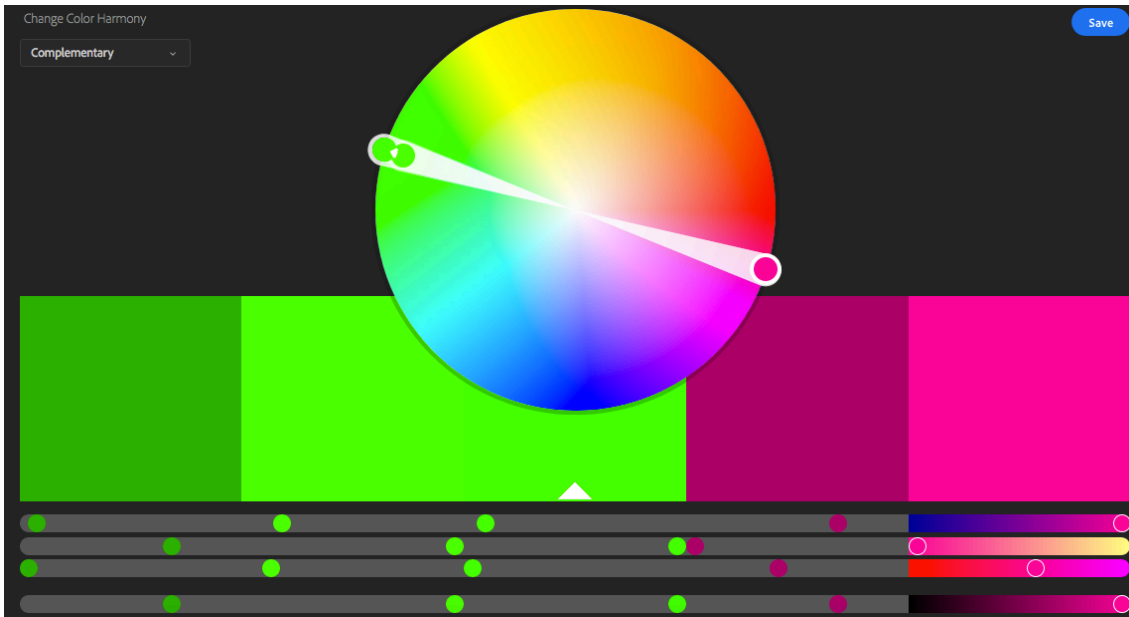


Figura 5. Ejemplo de aplicación de regla de complementación en el modelo HSL.

- La *triada de colores* es una combinación de tres colores espaciados equitativamente entre sí a lo largo de la rueda. Estos suponen un buen punto de comienzo en una serie de datos cualitativa e ir jugando con las sombras de cada uno de ellos.

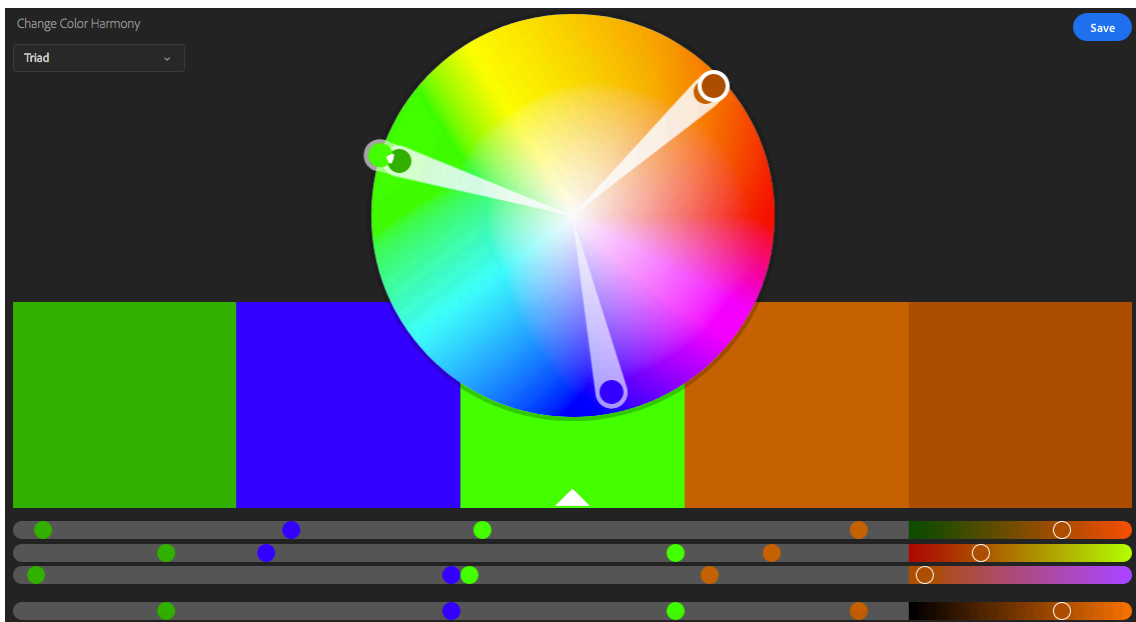


Figura 6. Ejemplo de aplicación de regla de triada en el modelo HSL.

Cabe decir que estas cuatro reglas no son ni mucho menos definitivas, pues para acercarse al caso de éxito que persigue una herramienta DataViz éstas se han de complementar con otras buenas pautas. Entre ellas, se encuentra por ejemplo la de no utilizar colores excesivamente saturados en una misma escala, pues podrían abrumar

demasiado el gráfico; o la de evitar colores con baja saturación y alto nivel de luminancia, ya que de esta forma no se apreciarían correctamente sin un fondo oscuro.

Además, es importante destacar en última instancia que existen ciertos tonos de color que ya están predefinidos o inculcados coherentemente en la sociedad. Un ejemplo de estos últimos podría surgir en la representación de series de datos del partido demócrata o republicano, ya que se podrían asociar los colores azul y rojo respectivamente.

### 2.3.4 Elementos gráficos en DataViz

Ahora que ya nos hemos familiarizado con los fundamentos de la percepción visual y una vez contempladas algunas de las nociones generales de diseño gráfico, se puede dar paso a la visualización de datos como disciplina y a los prototipos de herramientas que la componen.

Recordemos que la materia DataViz se encarga de representar ideas complejas llevadas a cabo de forma clara, precisa y eficiente [22]. Para ello, es necesario que los elementos gráficos presentes en cualquiera de sus herramientas posean las siguientes funcionalidades:

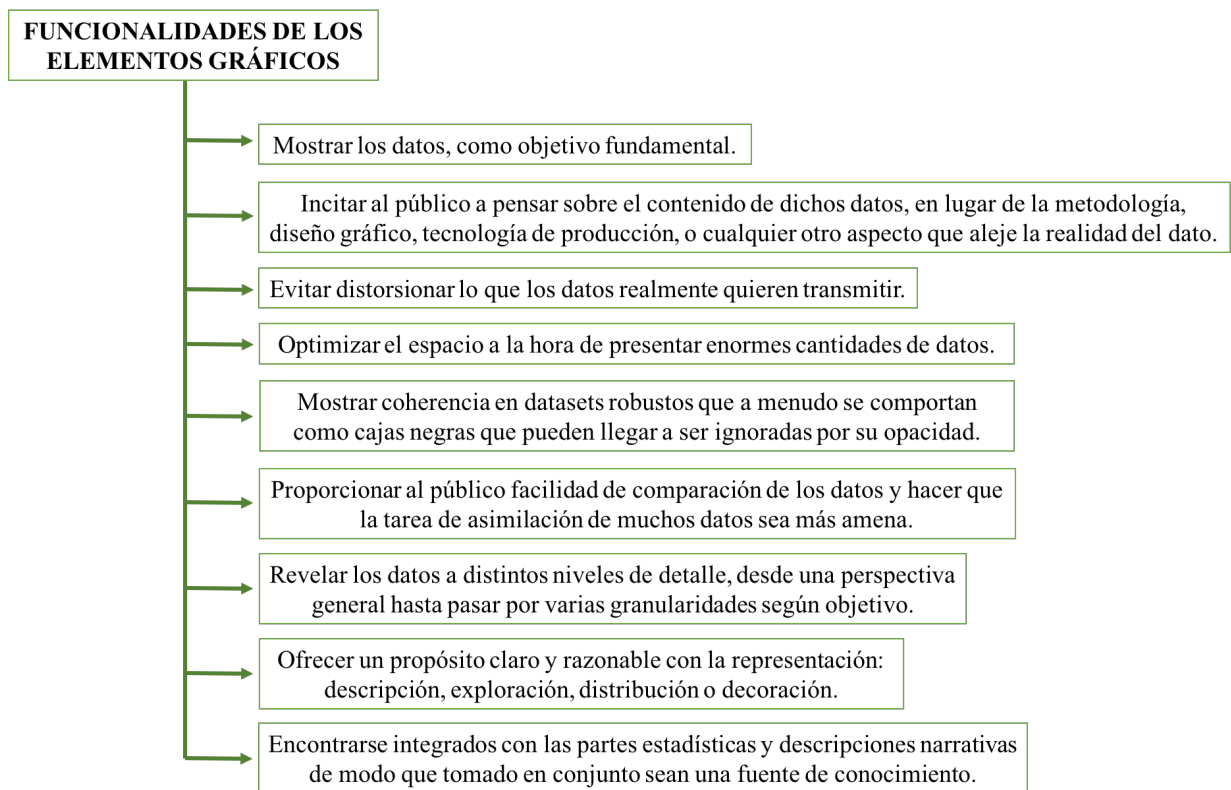


Figura 7. Esquema de funcionalidades de los elementos gráficos en DataViz.

De cualquier forma, todo se resumen a un mismo concepto: la transmisión de información, enfocada a todo usuario que busque cualquier método de razonamiento o conocimiento acerca del dato y a través de un conjunto de imágenes, texto y números.

Pues bien, la verdadera utilidad de las herramientas DataViz se tiene cuando estas tres partes (imágenes, texto y números) trabajan como un único mecanismo. Por ejemplo,

mostrando en un gráfico numéricamente los datos de los que el visualizador realmente quiere obtener ideas, aportando ciertos comentarios narrativos o ecuaciones a modo de justificación, integrando una leyenda para facilitar la identificación de series de datos, e incluso añadiendo pequeñas tablas con números con tal de ofrecer una mejor relación de la información.

Claro está que en la mayoría de ocasiones se dispone de demasiada información y se dedica poco tiempo a gestionarla, por lo que es complicado encontrar la forma de generar elementos visuales que proporcionen una idea clara de los datos y a la vez resulten atractivas para el usuario final. El conjunto de herramientas DataViz que se encarga de esto y por consiguiente de implementar todas las funcionalidades comentadas sobre la figura 7, se dividen en dos partes: la visualización de datos estática y dinámica.

Por un lado, la *visualización de datos estática* es aquella que muestra la información de forma plana, sin permitir al usuario interactuar con el material visual. Un ejemplo de este tipo de DataViz estática son las **infografías** [23], que tienen como objetivo principal convertir la transmisión de información compleja en una tarea sencilla, tratando de explicar esa información de la forma más amena y artística posible.

Estas infografías realmente no solo se muestran gráficos, sino también hechos, líneas de tiempo, historias, recetas, estudios de mercado, tendencias, encuestas, mapas, rutas, ubicaciones, biografías...

En la siguiente figura se muestra un ejemplo de infografía:



Figura 8. Ejemplo de infografía temporal.

Tal y como se puede observar, la infografía anterior representa los eventos que han ocurrido en la evolución del mítico personaje Mario Bros, ofreciendo una fácil comprensión sobre los distintos eventos históricos. No obstante, es una comunicación plana en la que el usuario no puede indagar sobre determinados puntos, y donde no se establecen variaciones en los datos.

No obstante, la *visualización de datos dinámica* se encarga de resolver el problema que deja el anterior, permitiendo al usuario final interactuar con los elementos visuales y navegar entre los datos para poder centrarse en los diferentes puntos de interés aplicando filtros, zoom, y otros muchos elementos de configuración y personalización.

Un ejemplo claro de este tipo de herramientas DataViz dinámicas son los **dashboards** [18], donde en una sola hoja se muestra una visualización gráfica orientada a los objetivos de negocio y monitorización de KPIs (Key Performance Indicator) con tal de optimizar las estrategias en las empresas.

A continuación, se representa un ejemplo de dashboard mediante la figura 9:

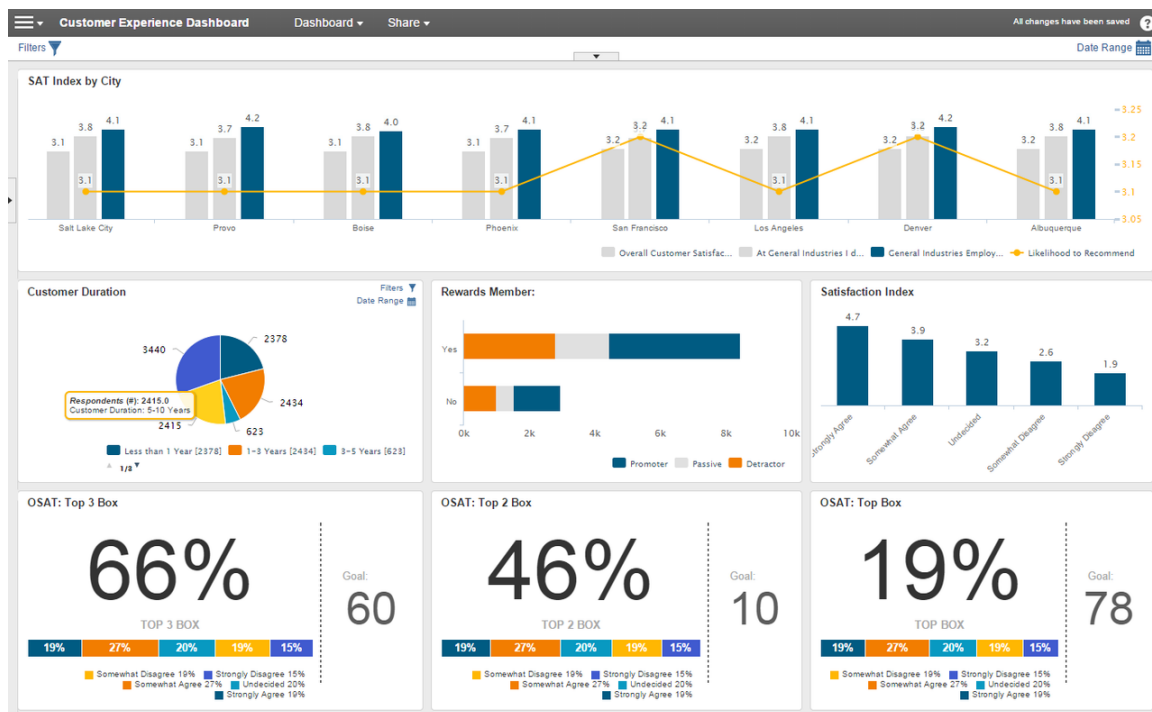


Figura 9. Ejemplo de dashboard orientado a experiencia de cliente.

En el anterior dashboard se observa el estado global de los principales indicadores de un negocio, concretamente los enfocados a la experiencia de cliente. Esto quiere decir que en caso de existir algún problema con los parámetros monitorizados se podría detectar y tomar decisiones al respecto; no obstante, en ningún momento se permitiría analizar las causas que originaron dicho problema.

### 2.3.5 *Storytelling: la importancia del contexto*

Es totalmente cierto que los tipos de elementos gráficos explicados anteriormente son altamente influyentes en la creación de herramientas DataViz. Pues bien, a todo esto cabe añadir un factor que en múltiples ocasiones se pasa por alto y tiene un elevado nivel de importancia en esta ciencia: el concepto de Storytelling, o lo que es lo mismo, el talento para contar historias [24].

Esto, en realidad, está ligado en cierto modo al apartado anterior, pues las historias son una forma maravillosa de enfocar la atención humana y ayudar a discernir por qué los datos presentados son importantes o relevantes en alguna parte de nuestra vida. En la siguiente figura [25] se puede observar la importancia del Storytelling en la ciencia DataViz:



Figura 10. Relación de Storytelling y DataViz.

Las capacidades que ofrece la técnica de Storytelling son las siguientes [26]:

- *Ilustrar un proceso* con la ayuda de herramientas DataViz simples y que el usuario considere fáciles de digerir.
- *Remarcar líneas de tendencia* a través de acontecimientos ocurridos, con el soporte de gráficos de línea, mapas, pie charts, etc...
- *Apoyar un argumento sólidamente* con gráficos personalizados que hagan al usuario interesarse por el contexto.
- *Conducir las emociones del usuario* presentando contenidos visuales de noticias que sigan el hilo y puedan provocar un impacto positivo en el público.
- *Crear interactividad* en las herramientas DataViz ofrecidas al contar la historia para proporcionar al usuario una experiencia personalizada.

Claro está que existen algunos conjuntos de datos que únicamente son significativos en un determinado contexto, y aquí es precisamente donde contar una historia a través de ellos se convierte en una acción de valor. Además, la parte del público receptora de

cualquiera de estas técnicas de Storytelling tiende a ubicar como punto central de la historia las percepciones gráficas sobre las que ha ido “viajando” en el transcurso de la misma; y esto es debido a que no se está meramente informando al usuario final, sino que se le está aportando entretenimiento de forma constructiva mediante la experiencia [27].

De este modo, la inteligencia de esta metodología se basa en conseguir una mezcla equilibrada entre informar y contar la experiencia, pues es lo que realmente se traduce como un caso de éxito en la comunicación. No obstante, no toda visualización requiere necesariamente una historia, pero sí una meta u objetivo específico en el cual es necesario dotar de un contexto.

## 2.4 **Ámbito de aplicación**

Ya se ha comentado que el presente Trabajo Final de Máster se enfoca primeramente desde una perspectiva genérica para posteriormente pasar a un caso práctico concreto usando datos de electromiografía. Esto quiere decir que en el próximo capítulo se estudia la metodología CBR que propone recuperar y adaptar visualizaciones en base a una caracterización sintáctica y semántica tanto de las características de los datos, la visualización y los requisitos del usuario; para en el capítulo 4 ser aplicada sobre un caso de estudio en particular y diseñar herramientas DataViz efectivas en plantillas de informes.

La coordinación de estas dos materias – CBR y DataViz – para la optimización de informes realmente podría orientarse realmente a cualquier de campo de aplicación donde se la generación de datos por sistemas IoT requieran de explicación visual:

- Niveles de contaminación en *Smart Cities*.
- Estado de congestión de las carreteras y otras *infraestructuras urbanísticas*.
- Niveles de consumo en la monitorización de maquinaria en *factorías*.
- Evolución de los niveles de actividad en *plantas de producción*.
- Medición de glucosa, fructosa, sacarosa y otros muchos factores influyentes en la monitorización de pacientes dentro del campo de la *Medicina*.

En nuestro caso, la iniciativa práctica se centra en la toma de respuestas neuromusculares de pacientes asociados a la Medicina del Deporte. Para ello, se parte de una serie de pacientes (o casos), a los que se les ha de asignar (como resolución al problema) una explicación visual personalizada y ajustada a las necesidades de los datos y herramientas DataViz acordes al contexto.

Llegados a este punto y hasta el final del presente capítulo se procede a entrar en el contexto que a posteriori va a suponer este caso práctico, por lo que se estudian las técnicas y procesos llevados a cabo en la colecta y tratamiento de datos en una población de pacientes.

### 2.4.1 *Análisis y representación de datos en el ámbito sanitario*

De forma general, el campo de la Medicina hoy en día es claramente *hot-topic* en materia de análisis y representación de datos, pues cada vez es mayor la variedad de dispositivos y sistemas encargados de recoger y tratar la información de los usuarios. Esto es, sin duda, consecuencia directa de que el mundo de Internet of Things se esté incorporando en este campo a una velocidad sin precedentes.

Por extensión, todo lo referente al ámbito específico de la Medicina de la Educación Física y el Deporte juega un papel clave en el análisis de datos de pacientes debido al tremendo alcance que posee. Dicho alcance se justifica debido a que, de forma general, los datos sanitarios no solamente se componen de la historia clínica, sino que también de cualquier otro recurso relativo a la salud. Esto se traduce en un extenso conjunto de datos que contienen información que, a día de hoy, resultaría impensable facilitar a desconocidos [28]. Entre ellos:

- Datos que recopilan información y valoraciones sobre la situación o evolución clínica de un deportista o paciente.
- Datos relativos al padecimiento (o riesgo) de lesiones, enfermedades o discapacidades.

#### ***2.4.2 La electromiografía como ciencia***

La electromiografía [29] es un procedimiento de diagnóstico que evalúa la condición física de los músculos y las neuronas motrices que los controlan. Por naturaleza, un movimiento neuromuscular consiste en la emisión de señales eléctricas que causan flexión y relajación de los músculos; por lo que un análisis EMG se encarga de traducir estas señales a modo de gráficos o series numéricas que ayudan a los doctores a llevar un diagnóstico.

Es así que los médicos (o cualquier experto otro sanitario asiduo) requieren de un análisis EMG para un paciente que tiene algún tipo de síntoma o desorden en las neuromuscular. Estos síntomas incluyen, entre otros: dolor, insensibilidad, debilidad en las extremidades e incluso hormigueos.

Para la realización de una prueba EMG se utilizan pequeños dispositivos llamados electrodos que actúan a modo de sensores recogiendo impulsos eléctricos en los movimientos de flexión y contracción de los músculos del paciente. Estos impulsos son los que el especialista requiere de análisis e interpretación para la detección de anomalías.

#### ***2.4.3 A quién se dirige***

Un análisis EMG, al igual que el campo de La Medicina del Deporte en general, está enfocado a [30]:

- Personas en busca de soluciones efectivas ante cualquier impedimento físico, enfermedad o discapacidad.
- Personas que se quieran iniciar en el deporte y quieran conocer su estado neuromuscular y su riesgo de padecer enfermedades.
- Personas con patologías traumatológicas del aparato locomotor que impidan el desgaste físico o incluso ciertos hábitos cotidianos.
- Personas que quieran mejorar su actividad física y les permita obtener un mejor performance en los entrenamientos.

#### 2.4.4 Responsables del análisis médico

Es así que existen una enorme cantidad de entidades por las que pueden pasar los conjuntos de datos desde que se obtienen en la realización de una prueba EMG hasta que se almacenan y se consultan a modo de representación visual. Dichas entidades engloban una serie de expertos que pertenecen a distintas áreas patrimoniales y funcionales [31]. De cualquiera forma, para el caso que nos ocupa se pueden listar de la siguiente forma:

- Hospitales y centros de salud.
- Clínicas especializadas y farmacias.
- Mutuas y compañías aseguradoras.
- Empresas dedicadas al desarrollo y gestión de herramientas de uso por profesionales de la salud.
- Empresas terceras contratadas con capacidad de acceso y almacenamiento de los datos.

#### 2.4.5 Alcance tecnológico del análisis

Así pues, todos los datos recogidos en las pruebas EMG son susceptibles de ser tratados, bien sea de forma escrita en un informe médico o automatizada mediante el soporte técnico más adecuado.

Desde un punto de vista tecnológico [32] estos conjuntos de datos podrían ser estructurados mediante un sistema de capas como se muestra en la siguiente tabla junto a un posible ejemplo práctico:

Capa	Contenido
Fuente de datos	Origen de la información recogida por los electrodos. Esto es, los impulsos eléctricos correspondientes a los movimientos flexión-relajación de la prueba EMG.
Integración	Adquisición e incorporación de los datos en las distintas plataformas de software del centro sanitario.
Storage	Almacenamiento de datos en sistemas Cloud o, como alternativa, en bases de datos de gran envergadura.
Computación Cloud	Manejo de datos del análisis EMG junto a otros recursos del historial clínico del paciente.
Representación y análisis	Visualización y generación de informes interactivos que permitan al experto focalizar en la detección de anomalías para un análisis EMG determinado en combinación con otros aspectos recuperados del historial clínico.

Tabla 3. Implementación tecnológica del análisis EMG.

Para comprobar la envergadura del anterior conjunto de capas y para finalizar la puesta en escena del estado del arte del contexto, se nombran algunas de las muchas aplicaciones prácticas en el campo de la medicina del Deporte y donde se aprecia considerablemente la presencia del análisis EMG como parte sustancial del IoT así como la necesidad de herramientas DataViz [33] para la optimización del análisis:

- Monitorización remota de los signos vitales de los deportistas, bien estén en situaciones atléticas normales, de riesgo extremo o bajo supervisión médica.
- Tracking del stock de medicación para deportistas o pacientes.
- Ubicación tanto de deportistas que hacen uso de wearables como de personas dependientes de profesionales.
- Control de la instrumentación de bombeado, paneles analíticos, camas de hospital, etc...

## 2.5 Recapitulación del estado de la materia

El presente capítulo es un precedente claro de cómo el CBR, como técnica de IA, es un sistema que se adecúa perfectamente a cualquier entorno donde existan conjuntos de datos masivos. Cabe añadir que, hoy día, es un hecho que estos conjuntos de datos procedan de numerosas fuentes de dispositivos IoT con un alto nivel de heterogeneidad y en muchas ocasiones totalmente opacos al usuario final, por lo que una propuesta de valor que los sistemas CBR pueden ofrecer en este aspecto es la visualización inteligente de estos datos.

Tal y como se ha visto, la ciencia DataViz supone un punto de inflexión en la capacidad de razonamiento humana, convirtiéndose la generación de explicaciones visuales en una forma de inteligencia. Si extrapolamos esto al ámbito sanitario, vemos que la incorporación de nuevos dispositivos IoT encargados de la recogida de datos en este entorno y toda la pila tecnológica que llevan de la mano hace que los expertos sientan real necesidad por disponer de informes donde se puedan interpretar estos datos de forma automática y con la mayor precisión posible, facilitando así la toma de decisiones en sus tareas.

En los siguientes capítulos entran en análisis las distintas fases para un sistema CBR genérico e independiente de dominio, de modo que pueda ser reutilizable en cualquier caso de uso que se requiera la recuperación de plantillas de informes donde primen la reutilización y adaptación de elementos gráficos.

Esto se lleva a cabo teniendo en cuenta en todo momento el alcance del proyecto global en el que se engloba el presente Trabajo Final de Máster, particularizando en el área de operación del mismo: la propia reutilización y adaptación de elementos DataViz.

A pesar de que en el transcurso del tercer capítulo todo esto se toma desde una perspectiva totalmente genérica, a posteriori – en el capítulo 4 – sí se particulariza dentro de este sistema CBR general un sub-sistema que pertenece al dominio específico de la Medicina del Deporte con datos EMG y que consiste en una herramienta de reutilización y adaptación de elementos gráficos.

## Capítulo 3. Metodología

Una vez analizado el estado de la materia que constata el presente Trabajo Final de Máster, se procede mediante este capítulo a describir la parte teórica que corresponde al diseño y desarrollo de un mecanismo inteligente de visualización basado en el paradigma CBR.

### 3.1 Alcance del desarrollo

Antes de nada, cabe decir que el presente Trabajo Final de Máster está ligado a un proyecto con una pretensión global que equivale a la metodología explicada en la sección 2.2:

1. Se dispone de una base de casos de plantillas.
2. Estas plantillas se caracterizan según los tipos de datos a visualizar y la función de visualización apropiada acorde al mensaje que se desea transmitir.
3. La información ha de ser etiquetada correctamente para adaptarla a un caso, que se corresponde con un tipo de plantilla.
4. La plantilla más adecuada se recupera para ser completada con las herramientas visuales reutilizadas y adaptadas a los datos previamente etiquetados.

Es por ello que estamos ante un sistema genérico e independiente del dominio que sigue una metodología CBR aplicable a cualquier caso de uso donde prime la recuperación de plantillas de informes. Esto supone lo que en adelante se denomina como *sistema CBR general*.

No obstante, el desarrollo que abarca este Trabajo no se corresponde con todo este ciclo CBR general, sino que toma una parte crucial dentro de éste. Esta parte se corresponde con las acciones de **reutilización y adaptación de las visualizaciones**, y supone un *sub-sistema CBR* dentro del ciclo general de la recuperación de plantillas de informes.

En el presente capítulo se parte de la comprensión de la base del sistema en su totalidad con tal de profundizar progresivamente en la metodología que ocupa este proyecto. Es precisamente por este motivo que ha de quedar clara la terminología que distingue el sistema CBR general del sub-sistema CBR propuesto para la reutilización y adaptación de elementos gráficos. Todo esto se abarca de forma genérica e independiente del dominio para otorgar de particularización al siguiente capítulo de la memoria.

### 3.2 Generación de plantillas de informes

Tal y como se ha comentado en el capítulo 2, a la hora de transmitir información, independientemente del medio que se trate, es imprescindible tener en cuenta tres necesidades básicas:

- La competencia *literaria*, que define la habilidad para pensar y comunicar con palabras, bien sea en lenguaje hablado o escrito.
- El ámbito *numérico*, que permite la transmisión de información mediante el alfabeto que comprenden los números.
- La parte *gráfica*, que habilita la comunicación mediante imágenes y todo tipo de representaciones visuales.

Para una comunicación efectiva estas tres competencias han de tratarse en total concordancia, y un sujeto perfecto para su utilización podría ser un informe correspondiente a cualquier tipo de área funcional [34]. En este caso, resulta más que factible narrar una historia con los datos y tener la capacidad de aportar valor a la vez que se transmite la información, con la finalidad de crear un cierto nivel de fidelización en el público al cual va dirigido.

Si un informe se divide de forma sustancial se pueden diferenciar claramente el contenido comentado en los tres puntos anteriores, teniendo la parte numérica como un punto de la narrativa que complementa la parte literaria y permite cuantificar los datos de forma alternativa a las palabras.

En la figura 11 queda representado un pequeño esquema:

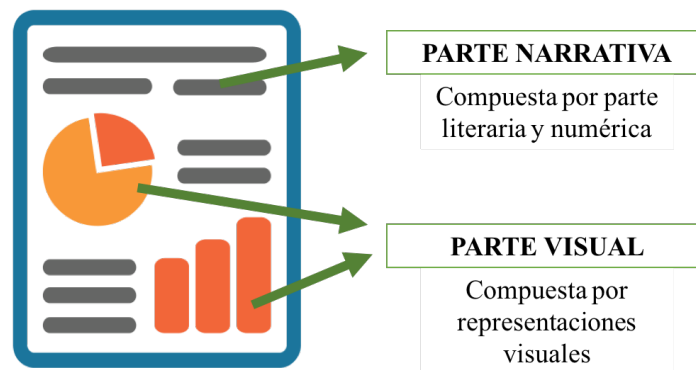


Figura 11. Partes sustanciales de un informe.

El uso de estos informes en nuestra metodología se basa un sistema CBR general donde los casos son las propias plantillas de informes. No obstante, además existe otro CBR que se utiliza para recuperar y adaptar los gráficos, es decir, son los elementos visuales los propios casos que se recuperan y adaptan para ser objeto de reutilización junto a elementos de texto.

Esto se muestra en la figura 12:

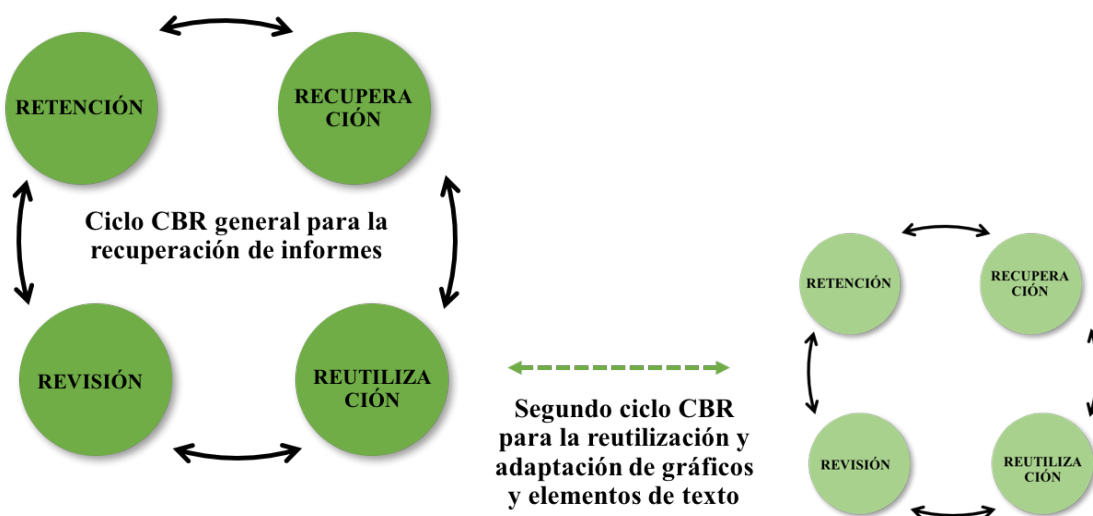


Figura 12. Esquemas de ciclo general para la recuperación de informes y ciclo secundario para la reutilización y adaptación de sus elementos sustanciales.

Ya se ha comentado en el transcurso de los anteriores apartados que en este proyecto se propone un enfoque CBR para aportar explicaciones visuales dentro de plantillas de informes y representar información de forma cuantitativa y cualitativa. Además, también se ha aclarado en el capítulo 1 que el objetivo específico de este Trabajo Final de Máster se centra en la parte visual de las explicaciones, y no en los textos, por lo cual el resto de documento se enfoca única y exclusivamente la información visual que ocupa dicha parte de los informes.

### 3.3 Sistemas de automatización de datos

Los sistemas CBR que se encargan de resolver la tarea principal de presentar datos en forma gráfica de manera automática ofrecen un conjunto de reglas que eligen por sí mismas entre distintos enfoques de representaciones visuales, lo cual hace que se simplifiquen notablemente las actividades de un programador de diseño de herramientas de DataViz evitando la necesidad de ajuste y representación de los datos cada vez que se actualiza un campo de salida dentro del conjunto.

#### 3.3.1 Ciclo CBR para la finalización de informes

El ciclo CBR general propuesto que se encarga de la recuperación de informes tiene como caso base el conjunto de plantillas cuyos huecos se pretenden rellenar con información visual. Así pues, tiene el aspecto que se indica en la siguiente figura:

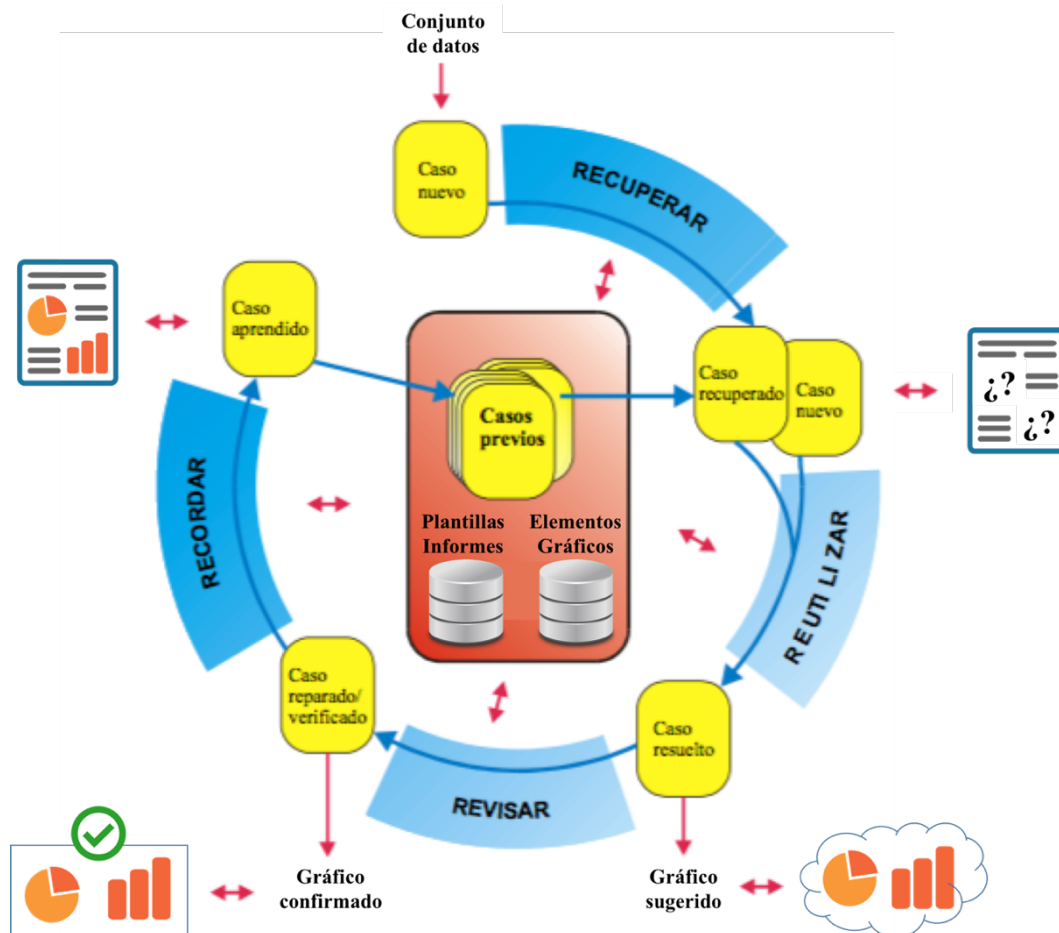


Figura 13. Ciclo CBR para la agregación de elementos gráficos en plantillas de informes [13].

Con respecto a este ciclo se destacan los siguientes puntos:

1. Primeramente, la petición de entrada es el propio material visual (chart), que a su vez viene acompañado de:
  - a. Conjunto de datos que se pretende agregar al informe.
  - b. El sujeto al cuál se dirige el mensaje del informe.
  - c. Las principales funciones de visualización.
2. En segundo lugar, la fase de recuperación de la plantilla de informe se basa en los modelos previos de plantillas de los que se dispone como caso base.
3. Seguidamente, la etapa de reutilización es la que comprende el segundo ciclo CBR para la recuperación y la adaptación de gráficos. Es decir, se consigue sustituir los huecos con los gráficos adecuados según la función de visualización y el conjunto de datos. Aquí se ha de tener en cuenta el tipo de gráfico apropiado a la función de visualización y a los datos de los que se disponen, así como los colores, formas y tamaños a utilizar para causar un buen impacto en el espectador. Los principios de Gestalt comentados en el capítulo anterior juegan un papel muy importante en la utilización de gráficos para captar la atención del usuario final. Cabe decir también que otro elemento de importancia en este segundo ciclo CBR que comprende la reutilización sería la adaptación de elementos textuales, lo cual queda fuera de los objetivos iniciales de este trabajo.
4. Una vez se recupera el gráfico, se sugieren los elementos visuales que mejor se adaptan a los datos de entrada y a los huecos del informe.
5. Así pues, a continuación se hace efectiva la fase de revisión para comprobar que los gráficos aportan valor al texto en el Storytelling, y por consiguiente que el informe es efectivo en su totalidad. Esto da paso a la siguiente etapa, que consiste en:
6. La interacción del usuario con los elementos visuales de dicho informe para posibilitar el aprendizaje, basándose en la observación y memorización de las configuraciones realizadas: vista general, zoom en las distintas partes, selección de rangos, exploración, extracción, filtrado, conexión...
7. Por último, en base a esta personalización, el sistema establece el caso aprendido eligiendo el valor inicial para las características visuales. Características como el tamaño de línea, color y grosor son retroalimentados de acuerdo a los datos y funciones de visualización tomados en el proceso. Estos cambios se almacenan en los ajustes del caso base.

### ***3.3.2 Jerarquía CBR para la finalización de informes***

Para diseñar los elementos de visualización acordes a este ciclo y cumplir la efectividad en la transmisión de información [35] es necesario disponer de una serie de criterios que se apoyen básicamente sobre:

- Las propiedades de la información que se pretende visualizar en base a los objetivos y las preferencias que el usuario persigue con la visualización.
- Las relaciones que tienen los datos entre sí y con el contexto, tratando la información dentro de un dominio específico y definiendo una taxonomía de análisis de datos que se pueda utilizar para caracterizar información heterogénea.
- Las técnicas visuales más apropiadas dentro de un determinado diseño, conociendo todos sus principios y sumando el valor que aporta la metodología del storytelling.

Un sistema muy apropiado para la generación y explicación de gráficos automáticos es el llamado SAGE [36], cuyo funcionamiento se basa en representar las principales características de los datos de un usuario considerando en todo momento las funcionalidades y estructuras soportadas por cada tipo de forma gráfica.

La arquitectura de dicho sistema se representa mediante la Figura 14:

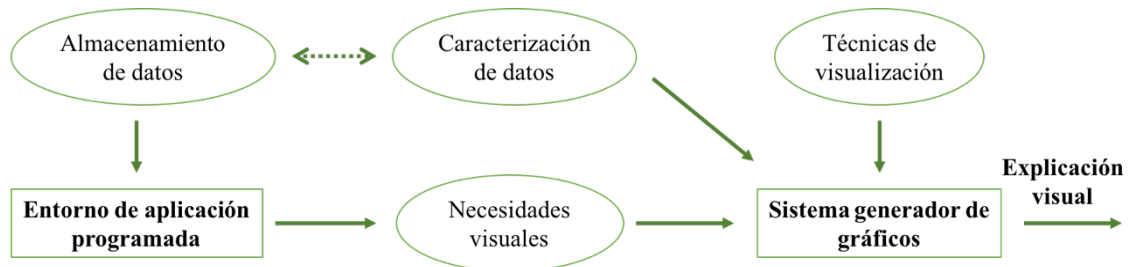


Figura 14. Arquitectura SAGE para la comunicación entre aplicaciones y sistemas gráficos.

En el anterior proceso se puede contemplar que el sistema SAGE se encarga principalmente de articular una aplicación con su correspondiente parte gráfica con tal de obtener una salida en forma de explicación visual.

La aplicación que se expone en la figura 14 puede ser cualquier interfaz de generación de informes encargada de realizar consultas al sistema Cloud o de bases de datos responsable de almacenar y estructurar la información según las entradas de los datasets. Una vez se obtiene esta información útil, la aplicación debe comunicar al sistema gráfico las necesidades de representación que requieren las salidas de los conjuntos de datos, de modo que finalmente se apliquen conocimientos de diseño DataViz para la selección y síntesis de técnicas gráficas apropiadas para una eficiente transmisión de las explicaciones al usuario final.

Si nos fijamos, esto es totalmente acorde al objetivo del sub-sistema CBR comentado, que se encarga de la reutilización y adaptación de elementos visuales. Es más, echando la vista a lo comentado en el anterior apartado de informes, vemos que esta metodología encaja perfectamente de la siguiente forma: en cada informe (correspondiente al ciclo CBR general) existen huecos para gráficos y, dependiendo de la descripción y particularización de dicho hueco, se pueden recuperar unos datos u otros expresados de forma gráfica (sub-ciclo CBR de reutilización y adaptación).

A continuación, se comentan las cuatro fases del proceso CBR como cuatro tareas estructuradas de forma jerárquica y que han de realizarse para cumplir con la tarea principal que propone SAGE, que no es otra que la óptima generación de explicaciones visuales.

### 3.3.2.1 Tarea de recuperación

La primera tarea de la jerarquía se divide en cuatro subtareas, mediante las cuales se identifican y procesan las entradas de datos, se asocian a casos anteriores parecidos, y finalmente se selecciona el óptimo. Se muestra la siguiente figura:

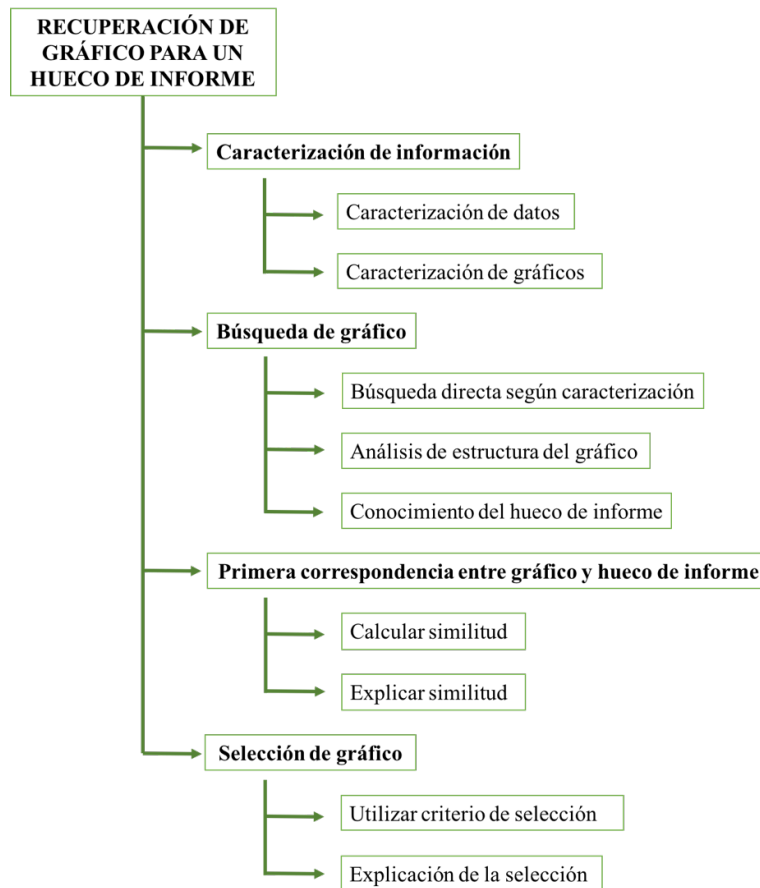


Figura 15. Jerarquía de subtareas para la recuperación de un gráfico en un hueco de informe.

Tal y como se observa, el conocimiento de los diseños no depende específicamente de la aplicación y por tanto el sistema no reconoce a priori de forma particular el tipo de datos ni las relaciones que existen entre ellos mientras están almacenados, sino que esto es necesario expresarlo en términos de *caracterización de la información*.

La caracterización de información permite entender las entradas de datos para un gráfico de una forma personalizada, lo cual hace que se entienda la elección del gráfico en base a un contexto determinado. Debido a la alta importancia que cobra esta subtarea, en los siguientes apartados se definirá y representará explícitamente la caracterización de datos y gráficos para que un sistema pueda tomar decisiones a la hora de presentar la información gráfica en plantillas de informes.

Una vez caracterizada la información, se tienen los gráficos candidatos a completar un hueco de la plantilla y se puede realizar un primer tanteo en el que se evalúe el acercamiento del gráfico al contexto de la información. Este acercamiento permite a posteriori elegir el gráfico correcto en función de los casos similares anteriores y la red semántica definida.

### 3.3.2.2 Tarea de reutilización

La reutilización del caso recuperado se enfoca desde dos perspectivas. Por un lado, las diferencias entre gráfico previamente seleccionado y el actual; y por otro, la adaptación de funcionalidades que el gráfico anterior puede ser transferido al nuevo gráfico. Se muestra en la figura 16:

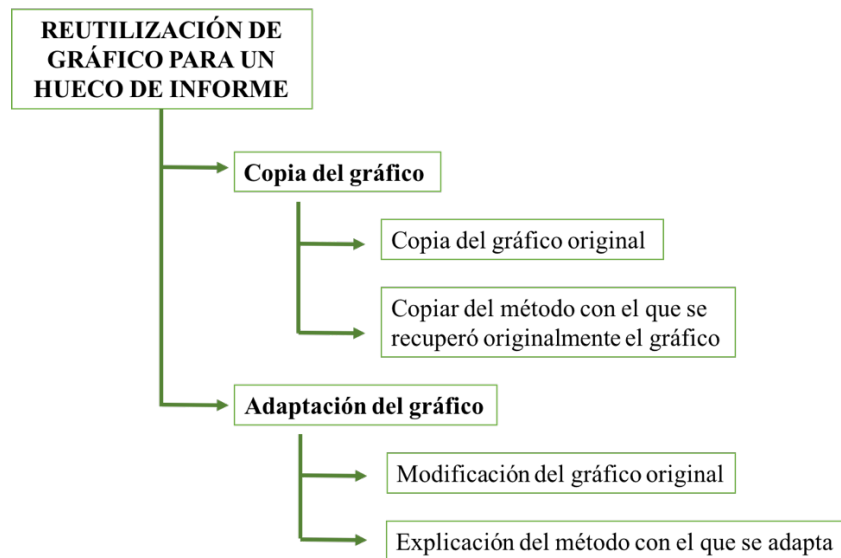


Figura 16. Jerarquía de subtareas para la reutilización de un gráfico en un hueco de informe.

La reutilización trivial de un gráfico es su propio duplicado; no obstante, la aplicación de este método normalmente no implica la resolución de la tarea ya que se necesita una adaptación entre el gráfico previo y el nuevo.

Por un lado, se contempla la equivalencia entre gráficos, por lo que se requiere una fuerte dependencia de las características contextuales con tal de generar los operadores transformacionales ligados a los atributos del nuevo gráfico.

Por otro lado, también se contempla la forma en la que se resolvió la tarea de recuperación en el gráfico que se duplica. Esto es, el gráfico original posee metadatos sobre la resolución de la tarea de recuperación sobre sí mismo, justificando el tipo de caracterización gráfica que se realizó, el tipo de informe para el cual estaba preparado, etc... Este gráfico original ahora duplicado es el más similar a al nuevo gráfico, y lo que se plantea es una serie de alternativas para el nuevo caso de contexto. Un ejemplo de esto podría ser la modificación de los colores para el nuevo gráfico en caso de que se quieran representar unas características acordes a una paleta determinada.

### 3.3.2.3 Tarea de revisión

Cuando un gráfico reutilizado no es el correcto, la metodología CBR sugiere una oportunidad de aprender del fallo generado.

Para completar esta tarea se sigue la jerarquía mostrada en la figura 17:

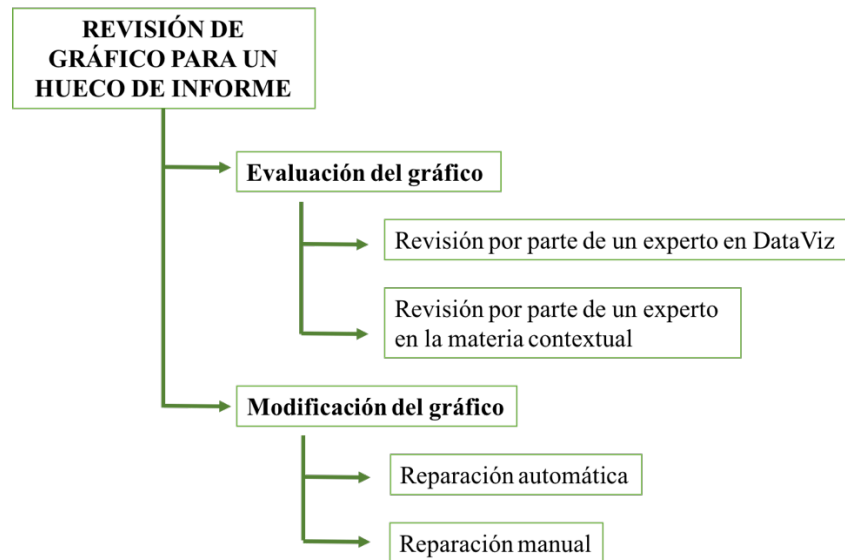


Figura 17. Jerarquía de subtareas para la revisión de un gráfico en un hueco de informe.

En primer lugar, se evalúa el gráfico dotándolo de opción de consulta; por un lado, a un experto en DataViz para comprobar la reutilización de los atributos gráficos; y por otro, a un experto en la materia contextual que sea capaz de asegurar si el gráfico elegido realmente proporciona conocimiento sobre los conjuntos de datos que originalmente se introdujeron en el sistema CBR. Al requerir de operación externa al sistema, esta tarea implica que nuestra metodología tome tiempo para la evaluación del caso.

En caso de que el gráfico necesite ser modificado el sistema aprende la solución tomada para revisar de nuevo que se encuentra ante el gráfico adecuado al hueco del informe. No obstante, si el fallo persiste, se puede tomar acción manual para la personalización de los atributos visuales a los datos de entrada.

#### 3.3.2.4 Tarea de retención

Tras la correcta personalización del gráfico para las entradas de datos propuestas, la última tarea del proceso se basa en la retención de la información por parte del sistema para incorporarlo en su base de conocimiento. El aprendizaje de acierto o fallo en la personalización del gráfico viene determinado por el proceso de evaluación y posible reparación del gráfico, pues en base a esto se determina qué información del gráfico hay que retener, en qué forma se debe hacerlo, cómo indexar este gráfico para otros casos de informes similares y cómo integrar el nuevo caso en la estructura de la memoria del sistema.

Para completar esta tarea se sigue la jerarquía mostrada en la figura 18:

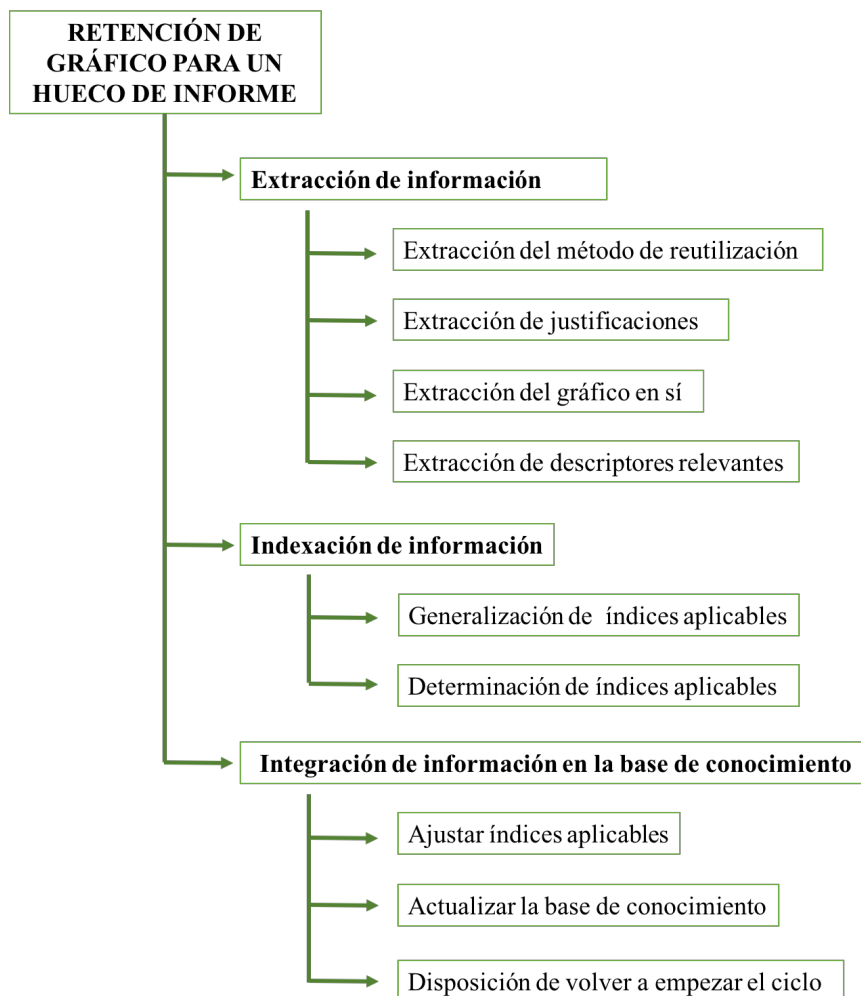


Figura 18. Jerarquía de subtareas para la retención de un gráfico en un hueco de informe.

La Base de Casos de elementos visuales para la complementación de plantillas de informes se actualiza independientemente de la forma en la que se elija el gráfico. De modo que:

- Si se resuelve mediante el uso de un gráfico anterior, se podría construir uno nuevo a partir del caso similar, tal y como se ha comentado previamente.
- Si el gráfico se construye por otros medios (como por ejemplo con la ayuda externa de un usuario), de forma irrevocable habría que generar un nuevo gráfico de forma completa.

De cualquier manera, además del propio gráfico se ha de extraer siempre una explicación u otra forma de justificación de la personalización del mismo, así como otros valores descriptivos relevantes. Sin ir más lejos, un buen hábito en la retención de información es la diferenciación de los casos de fallo en la tarea de revisión y la asociación entre ellos. Esto permite buscar patrones y establecer recordatorios en el sistema para en la reutilización de casos similares tomar acciones apropiadas, así como entender y corregir de forma inmediata fallos parecidos en la tarea de revisión.

Así pues, una vez se extrae la cantidad de información adecuada, se indexa de forma directa en función de la caracterización de datos de entrada; y de forma indirecta mediante las justificaciones tomadas en el proceso. Esto quiere decir que las características primarias indexadas son en realidad la propia caracterización de datos y

gráficos, y cuando entra un nuevo tipo de datos éstas se reajustan para mantener un mismo patrón en la memoria de casos. De otra forma, las características asociadas a las primarias son las indexadas como secundarias y tienen información relativa a las acciones tomadas en el proceso.

Un ejemplo de ajuste de índices puede resultar efectivo con el uso de una matriz capaz de relacionar características sintácticas con características semánticas, para así asignar unos pesos a cada relación y que estos pesos se vayan actualizando en función del feedback por parte de expertos en el contexto. Esta valoración indicaría el acierto o fallo de la personalización del gráfico y por tanto el reajuste de los índices primarios.

Con todo esto, la tarea principal de personalización de gráficos para la complementación de informes en un proceso CBR termina con la subtarea de integración y actualización de la base de conocimiento del sistema, donde se queda a la espera de la introducción de un nuevo conjunto de datos a representar.

### 3.4 Caracterización de datos

Se ha visto que la caracterización de datos supone una tarea imprescindible en el ciclo CBR general de la recuperación de plantillas de informes. Esta caracterización de datos persigue primordialmente el reconocimiento y la clasificación de nuevas entradas de datos para aplicar algoritmos y técnicas de captura de características, patrones, predicciones, anomalías y dependencias con otros tipos de variables.

Así pues, aquí entran en juego diversas técnicas que engloban desde procesos de filtrado y limpieza de datos hasta algoritmos de analítica avanzada. Esto permite, entre otros aspectos, el estudio de los conjuntos de datos y su evolución en distintos periodos temporales, establecer patrones y buscar tendencias inesperadas; y cómo no, analizar cómo se relacionan los datos con el contexto.

Toda caracterización efectiva de datos [37] ha de presentar un enfoque lo suficientemente flexible para acomodar nuevos tipos de datos, así como un alto nivel de comprensibilidad para capturar la relación entre la información y el entorno gráfico apropiado. Los puntos para proceder a explicar la caracterización de datos son:

- Análisis de las características utilizadas.
- Relación entre los conjuntos de información de carácter heterogéneo, tanto cuantitativa como cualitativa, en entornos dinámicos y estáticos.
- Características de datos que se han de mapear con técnicas gráficas.

De forma general, el sistema SAGE comentado anteriormente sobre la figura 14 divide en tres las dimensiones utilizadas para describir las características de los datos. Se esquematizan a continuación:

Dimensión	Definición
Orden	Orden en el que se exponen los datos. Bien sea cuantitativamente, ordinalmente e incluso sin seguir un orden determinado
Tipo	Si se dispone de factores de medida o cantidades de datos en función de unas coordenadas.
Dominio	Categorías conocidas como el tiempo, espacio, masa...

Tabla 4. Categorización de las características de los datos impuestas en el sistema de automatización SAGE.

Las entradas de datos se categorizan según las anteriores dimensiones, y para cada elemento de los conjuntos se tiende a observar unas determinadas relaciones o reglas de mapeado, que se basan en las siguientes premisas:

- Qué elemento de un conjunto de datos tiene al menos otro elemento en común con otro conjunto. Es decir, la búsqueda de relaciones como tal.
- La cantidad de elementos que pueden ser mapeados desde un conjunto de datos de procedencia a otro, como bien pueden ser el conjunto de características básicas entre dos datos con la misma estructura.
- Si existe una exclusividad para el mapeado, en caso de que los elementos de un conjunto de datos estén estrictamente ligados a otros, como bien puede ser un número identificador.

En el caso que nos ocupa, para la reutilización y adaptación de gráficos en un hueco de informe se han de establecer una serie de etiquetas y concebir un lenguaje común entre los datos cuantitativos y el contexto.

### 3.4.1 *Etiquetas sintácticas*

De forma general, las etiquetas sintácticas son parámetros que encargan de describir cálculos numéricos sobre una *población*, de modo que se haga referencia a algunos de sus aspectos estadísticos y probabilísticos que influyen sobre ella. Así pues, la utilidad del dato desde el punto de vista matemático implica que exista cierta distinción entre la ciencia de la estadística y la probabilidad [38]. Esto se justifica ya que la estadística ofrece comprensión y análisis, por un lado; mientras que la probabilidad ofrece una predicción y unas posibles consecuencias en base a modelos aleatorios.

De cualquier forma, se estudian a continuación ambos tipos de etiquetado sintáctico.

#### 3.4.1.1 *Etiquetas estadísticas*

Para comenzar nos centramos en la perspectiva estadística, que permiten entender los datos numéricos (o parámetros) presentados de una forma ordenada y, en cierto modo, sistemática. Estos conjuntos de datos estadísticos se dividen en cuatro grupos [39].

Por un lado, las etiquetas sintácticas referentes a la *posición de los datos* dentro de la población dividen un conjunto de datos en grupos con la misma cantidad de muestras. Se debe destacar el término:

- **Cuartil:** Valor que establece un cuarto de la distribución de datos. En toda muestra existen tres cuartiles; de modo que el primero deja un cuarto de distribución por debajo, y el tercero deja un cuarto por encima. El segundo, por tanto, es precisamente el que establece la mitad de la muestra (mediana) dentro de esas cuatro partes proporcionalmente iguales.
- **Percentil:** Valor de variable por debajo del cual se encuentra el porcentaje de una frecuencia acumulada dentro de la muestra. Por ejemplo, el percentil de orden 30 deja por debajo al 30% de los datos de una muestra, mientras que el 70% quedaría por encima.

En segundo lugar, las medidas de *centralización* indican valores con respecto a los que los datos parecen agruparse. Éstas son:

- **Valor medio:** Describe la tendencia central de los datos en base a todas las observaciones. Ésta puede verse afectada por los picos de distribución.
- **Mediana:** Hace referencia al valor que divide las observaciones en dos grupos con el mismo número de muestras. Es conveniente utilizarlo cuando los datos no son asimétricos. Equivale al percentil 50.
- **Moda:** Equivale a los valores donde la distribución de frecuencia de los datos alcanza su máximo.

Así pues, las etiquetas que marcan la *dispersión* de una muestra son las que indican la mayor o menor concentración de los datos con respecto a las anteriores medidas de centralización.

- **Picos de distribución:** Son los valores críticos dentro de la muestra, tanto por la parte que respecta a valores elevados como a valores bajos. De cualquier forma, ambos permiten identificar un posible valor atípico o un error en los inputs.
- **Rangos:** Se trata de la diferencia entre los valores mayores y menores dentro de una misma población.
- **Varianza:** Mide la dispersión alrededor de la media. Cuanto más pequeña sea, quiere decir que más concentrados se encuentran los puntos alrededor de dicho valor promedio. Sus unidades son el cuadrado de las de la variable, por tanto es muy sensible a valores extremos, que se alejan de la media notablemente.
- **Desviación típica:** Es la raíz cuadrada de la varianza, y por ello posee la misma dimensionalidad que la variable. Cabe destacar dos factores con respecto a esto:
  - Centrando en la media y a una unidad de desviación típica de distancia, se tiene más de la mitad (68%) de las observaciones.
  - A dos unidades de desviación típica con respecto a la media se tiene casi todas las muestras (95%).

Por último, las etiquetas sintácticas de *distribución* se encargan de informar sobre la forma en la que se organizan los datos dentro de la población. Aquí destacan:

- **Coefficiente de simetría:** Se encarga de medir si la muestra de datos se distribuye de igual forma en ambos lados de la media.
  - Si el coeficiente es negativo, la dispersión se vuelca mayoritariamente hacia la izquierda de la media.
  - Si el coeficiente es cero, la muestra es simétrica alrededor de la media.
  - Si el coeficiente es positivo, la dispersión se vuelca mayoritariamente hacia la derecha de la media.

### 3.4.1.2 Etiquetas probabilísticas

Distinguiendo de lo anterior, las etiquetas probabilísticas comprenden el conjunto de datos numéricos que, además de haber realizado cálculos sobre ellos, tienen un determinado grado de aleatoriedad.

Como las etiquetas sintácticas se particularizan por tener varios tipos de variables independientes, es completamente necesario determinar los modelos probabilísticos más utilizados [40] en la relación de dichas variables. Para ello nos vamos a basar en el diagrama de dispersión donde cada punto hace referencia a una muestra, y donde se pretende reconocer si existe relación entre las variables, de qué tipo es ésta y la posibilidad de predecir el valor de una de ellas en función de la otra [41].

Por un lado, la **correlación** comprende el grado de asociación entre dos variables [42]. Se expresa mediante un coeficiente (r), que se define según la ecuación 3.2:

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (3.1)$$

El rango de valores de r oscila entre -1 y 1, de modo que:

- Si  $r=1$ , existe una correlación positiva perfecta; es decir, una relación directa entre las variables. Cuando una aumenta, la otra también lo hace de forma proporcional.
- Si  $0 < r < 1$ , existe una correlación positiva.
- Si  $r=0$ , no existe relación lineal, por lo que habría que considerar otros métodos de interpretar el diagrama de dispersión.
- Si  $-1 < r < 0$ , existe una correlación negativa.
- Si  $r=-1$ , existe una correlación negativa perfecta, mediante la cual una variable indica una dependencia total sobre la otra. Esto es, los cambios en una variable implican un cambio directamente proporcional en la otra variable.

Así pues, la técnica de correlación no se trata de forma aislada ya que existe un método estrechamente ligado a éste y, de hecho, ambos son considerados como dos aspectos de un mismo problema. Este método no es otro que la regresión.

La **regresión lineal** es una técnica que cuantifica la relación en un diagrama de puntos dispersos correspondientes a dos variables con tendencia rectilínea. El ajuste de la relación sigue la ecuación:

$$y = a + bx \quad (3.2)$$

Donde “y” es la variable dependiente representando el eje vertical del gráfico y “x” es independiente mostrando la coordenada horizontal. No obstante, la ecuación 3.1 puede generalizarse para el caso en el que haya más de una variable independiente. Esto se denomina regresión lineal múltiple, y cada una de ellas tendría un coeficiente de regresión (coeficiente “b”), que indica la pendiente de la recta en el diagrama de dispersión; así como un intercepto (coeficiente “a”), que puede ser negativo, positivo, o cero. A partir de estos valores es posible trazar a lo largo de los puntos dispersos de un gráfico la línea recta que mejor se ajusta a cada una de las variables independientes.

Una de las formas más eficientes de concretar la bondad de la estimación de la regresión lineal es mediante el método de los mínimos cuadrados, que consiste en calcular la suma de las distancias al cuadrado entre los puntos reales y los definidos por la recta estimada; de modo que la mejor estimación es la que minimiza estas distancias.

Por otro lado, la **regresión logística** [43] se utiliza para determinar cómo una serie de factores influyen en una categoría dicotómica; o lo que es lo mismo, comparar los sucesos en base a dos posibilidades. Para ello, se tiene una o más variables independientes ( $x_i$ ) y una variable dependiente dicotómica (y) que informan sobre la probabilidad final.

Para poder interpretar el resultado del modelo de regresión logística se debe recurrir al concepto de “odds”, ya que es la probabilidad de presentar una característica o no presentarla, medida según el cociente de la ecuación 3.3:

$$odds = \frac{p}{1 - p} \quad (3.3)$$

Esta medida es muy flexible ya que se pueden comparar dos odds en forma de ratio para conocer si las muestras que poseen (o no) una característica determinada, también (o tampoco) poseen otra característica. Esta razón de productos cruzados es lo que se denomina el odds-ratio (OR).

Para el caso de la regresión logística, la opción más comúnmente usada para comprobar la efectividad es el método de máxima verosimilitud, mediante el cual básicamente se mide a partir de los datos de la muestra cuánto más probable es un modelo frente al otro, mediante la diferencia de los cocientes de verosimilitud de dichos modelos.

### 3.4.2 Etiquetas semánticas

Las etiquetas semánticas componen la información sobre el significado de los datos. Es decir, traducen las anteriores etiquetas sintácticas en un lenguaje apropiado al contexto en el que se toman. Este tipo de etiquetas se encuentran estrechamente ligadas al término de *Thick Data* [44], que defiende la idea de que los números no son totalmente suficientes para una interpretación real de cara a un usuario, por lo que se han de buscar una terminología personalizada en base al contexto.

Es por ello que las etiquetas semánticas, en conjunto con las etiquetas sintácticas descritas anteriormente, tienen un potencial tremendamente valioso.

- Por un lado, permiten crear un contexto y obtener conocimiento desde el punto de vista de las etiquetas sintácticas de valor estadístico, mediante el análisis exhaustivo de los conjuntos de datos y la detección de anomalías en ellos.
- Por otro lado, dotan de valor a las etiquetas sintácticas probabilísticas al permitir al usuario final la realización de estimaciones y comprensión de lo que en ocasiones se puede intuir de manera errónea.

Se puede afirmar que las etiquetas semánticas, por tanto, hacen que nuestros conjuntos de datos vayan más allá, que tengan sentido y sean capaces de completar los huecos vacíos que las etiquetas sintácticas dejan en un contexto. Al fin y al cabo, no tiene sentido visualizar datos de forma cuantitativa si se carece de una parte cualitativa capaz de proporcionar dimensiones optimizadas a un entorno particular [45].

El manejo de estas etiquetas resulta totalmente clave en cualquier aplicación del campo de la investigación. Mediante ellas se proporciona conocimiento, por ejemplo, sobre las razones y/o motivos por las que una persona sufre una enfermedad; o bien las características sociales que hacen que esta persona se relacione con otro prototipo de paciente que sufre otra enfermedad asidua. 1

## 3.5 Caracterización de gráficos

Una vez conocida la estructura interna de los datos se puede dar paso a la caracterización del material visual que complementa las etiquetas previamente descritas. Tal y como se ha comentado en el segundo capítulo, un gráfico construido y elegido correctamente implica una gran facilidad para contar una historia de un solo vistazo; no obstante, cabe decir que esta tarea no siempre resulta ni mucho menos fácil. La

visualización de información siempre se encuentra ligada a un objetivo específico, y esto depende estrictamente del ámbito en el que se encuentren los conjuntos de datos en cuestión.

De forma general [46], una herramienta DataViz se considera efectiva cuando cumpla las siguiente premisas de cara al público:

- Facilitar la coherencia de la información para extraer conocimiento del conjunto de datos en cuestión.
- Incorporar mecanismos de percepción humana en sus técnicas de visualización para una optimización de las funcionalidades.
- Capacitar la presentación de la información en base a su contexto. Esto es, relacionar etiquetas sintácticas y semánticas.
- Aumentar el nivel de claridad de la exposición a la vez que disminuye el tiempo requerido para dar valor a los datos.

Para el cumplimiento de las mismas, en este trabajo se proponen varios tipos de funciones de visualización para comunicar los datos en un informe, con diversos tipos de gráficos apropiados al mensaje que se pretende transmitir. Esta metodología se esquematiza mediante la siguiente figura:

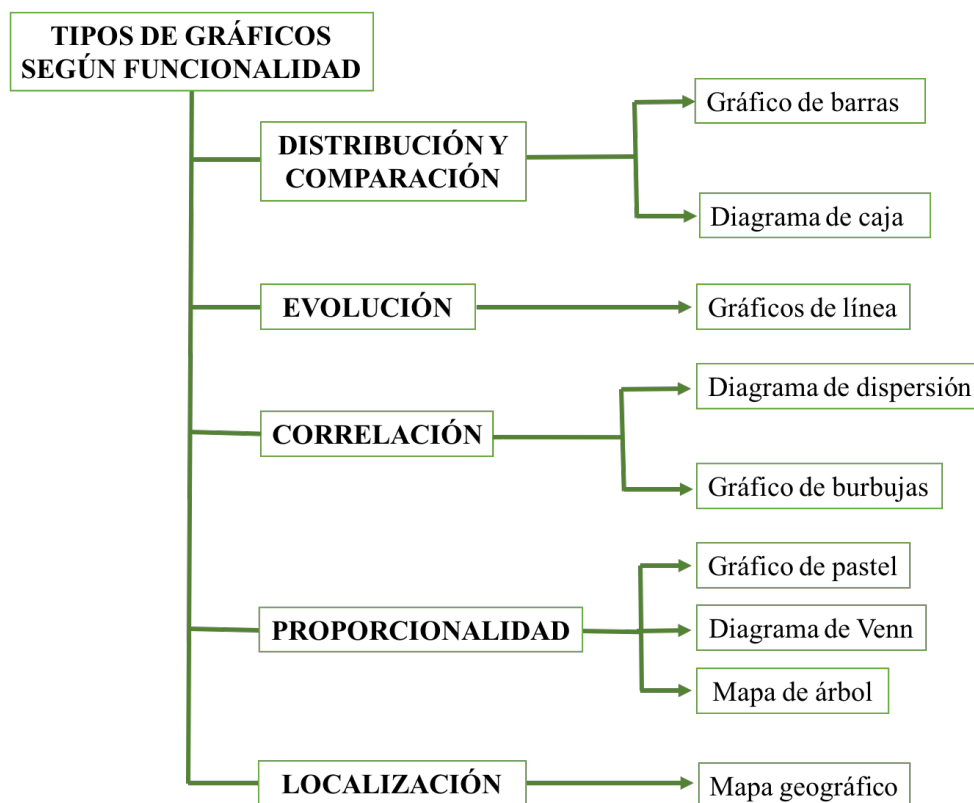


Figura 19. Esquema de tipos de funciones de visualización a transmitir en un informe y gráficos asociados.

El mensaje a transmitir es totalmente dependiente de la caracterización de estos gráficos, y de forma directa esto se puede enlazar con la *capacidad de percepción humana* y la *teoría del color* analizados en el apartado 2.4, pues el don de adaptación de un gráfico se basa en la estrecha relación que guardan estos conceptos. Entre las reglas

generales más comunes para la correcta reutilización genérica de los elementos visuales se destacan las siguientes premisas:

- Las líneas de ejes han de tener un *grosor* mínimo pero visible, en colores oscuros y donde se indique la granularidad adecuada a los valores de la métrica o dimensión a mostrar en el mismo.
- Además, la *textura* que conforme el fondo de 2 ejes tiene que estar compuesto por líneas de color también oscuro pero distinguibles totalmente de los ejes, creando una rejilla totalmente simétrica y con *orientación* vertical y horizontal.
- En caso de querer indicar cálculos estadísticos y/o probabilísticos se debe suprimir la textura del fondo del gráfico e insertar *modelos de líneas* señalando los valores destacables, eso sí, con un grosor superior al de los ejes e incluso jugando con la separación discontinua de su forma.
- También los elementos de las *conexiones y las cajas* se han de distinguir de los ejes y otras líneas estadísticas, pudiendo aumentar el grosor en los bigotes para destacar valores atípicos.
- El *tamaño y posición* de las formas que representan valores útiles cuantitativos ha de ser el suficiente para distinguir las categorías, pero teniendo en cuenta el número de formas que se incluyen en el gráfico con tal de que no se solapen entre sí.

Una vez visto esto, a continuación se da paso a la explicación de los principales tipos de material visual nombrados en la figura 19 y que se pueden utilizar en nuestras técnicas genéricas de adaptación de gráficos. Cabe decir que en este proyecto se utilizan las funciones de visualización proporcionadas en *The Data Visualization Catalogue* [47] pues ofrece una lista de gráficos categorizados por sus funcionalidades en cuanto a representación y tipo de mensaje que se pretende transmitir a la audiencia con dicho gráfico.

Cabe destacar que la elección de este catálogo como referencia no supone un patrón definitivo, pues no existe ninguna guía que establezcan la relación entre los gráficos y sus funcionalidades a modo de última instancia. No obstante, bien es cierto que los métodos sugeridos en el catálogo mencionado se adaptan muy adecuadamente a las necesidades de una comunicación. Es por ello que seguidamente se explican y ejemplifican los modelos de anatomía de cada uno de ellos.

### **3.5.1 Gráficos de barras**

En primer lugar, los *diagramas de barras* proporcionan cierta versatilidad en comparaciones nominales gracias a la longitud de las barras, siguiendo un orden entre distintas categorías. Un eje del gráfico distingue las distintas categorías que se comparan, mientras que el otro muestra los valores en formato discreto. A pesar de poder tener un aspecto similar a un *histograma*, se distinguen claramente ya que éstos muestran no muestran información categórica sino sólo la evolución continua de una única variable numérica.

Esta comparación se muestra con sus respectivas anatomías, en la figura 20:

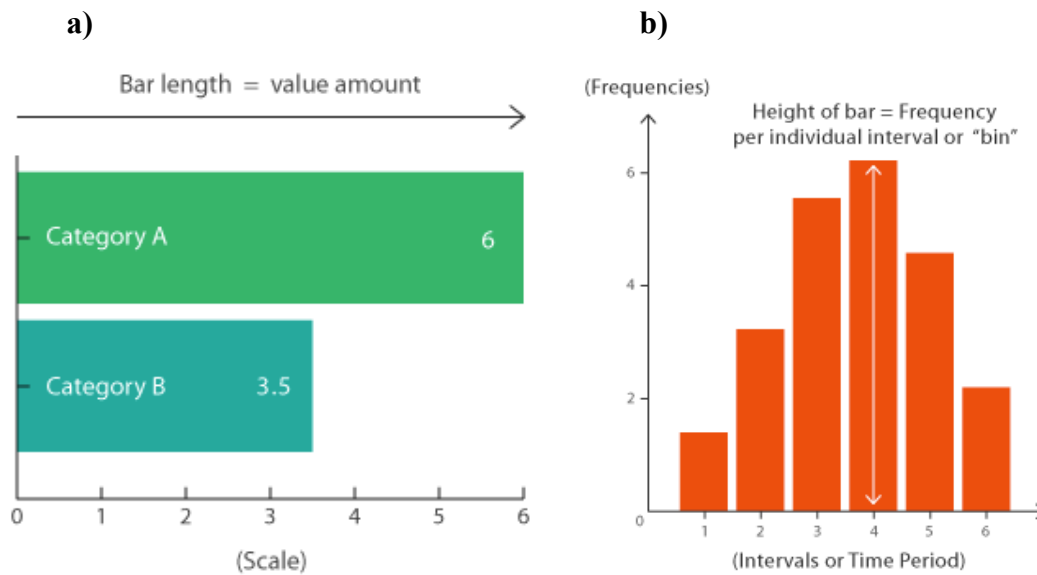


Figura 20. Anatomía de un gráfico de barras (a) y un histograma (b).

De esta forma, si esto se orienta por ejemplo de cara a un control de consumo alimenticio, el diagrama de barras de la figura 20 a podría representar la cantidad de dos categorías de comida consumidas, por ejemplo, *frutas* y *lácteos*, distinguiendo con colores de un mismo tono jugando con la saturación y luminancia. Por otro lado, el histograma de la figura 20 b sería totalmente influenciado por la variable temporal y resultaría adecuado para contemplar cómo un segmento de la población consume, en cantidades, un tipo de alimentos con el paso del tiempo, para lo cual se podría utilizar un único tipo de color.

Entre algunos errores comunes en este tipo de gráficos destacan la representación del valor promedio de una categoría, el hecho de utilizar demasiadas categorías, y por último, no mantener un orden de los valores de mayor a menor.

### 3.5.2 Gráficos de líneas

Por su parte, los diagramas de línea se utilizan para representar valores cuantitativos sobre un determinado intervalo, y es por ello que normalmente muestran tendencias y analizan cómo los datos evolucionan, principalmente, en una serie temporal.

Para ello, se parte de un sistema de coordenadas cartesiano donde el eje Y posee los valores cuantitativos y el eje X representa la escala de tiempo. La dirección de las líneas del gráfico es el propio indicador de tendencia, de modo que una pendiente positiva significa un incremento, y una pendiente negativa un decremento de los valores cuantitativos. Esto facilita la interpretación de tendencias entre distintos patrones y su similitud en la línea del tiempo.

Así pues, si en un mismo gráfico se agrupan distintas *líneas planas*, se pueden comparar entre sí diferentes series de datos y por consiguiente conseguir un contraste entre sus tendencias. Por otro lado, si lo que se quiere es ver el desglose de las series de datos dentro de una misma variable, es muy recomendable representar cada una de las líneas con un área de color en el fondo de modo que las series de datos queden *apiladas entre sí* distinguiendo claramente el desglose con respecto a dicha variable.

La comparación de estas dos funcionalidades se muestra mediante la siguiente figura:

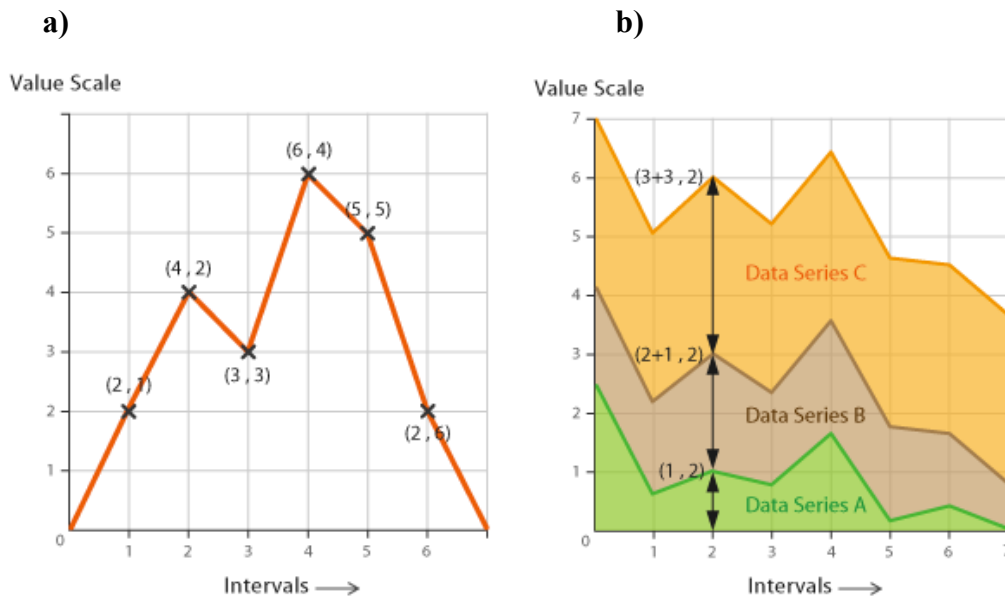


Figura 21. Anatomía de un gráfico de línea plano (a) en comparación con uno de áreas apiladas (b).

De esta forma, si por ejemplo se sigue el control de un paciente en sus niveles de glucosa, el diagrama de líneas de la figura 21 a podría representar con exactitud qué niveles se tiene remarcando incluso dónde se encuentran los niveles máximos y qué valores típicos de población se tienen. Por otro lado, las líneas apiladas de la figura 21 b resultarían útiles para destacar los niveles de un par de pacientes utilizando tonos de color completamente distintos, y así poder hacer una comparación además de una evolución, permitiendo incluso obtener patrones de los niveles de glucosa si los pacientes se caracterizan por tener cierta similitud en el contexto.

Cabe destacar que, tanto para el gráfico de líneas plano como apilado, existe un error común: la representación de demasiadas líneas de datos por gráfico. La utilización de más de 3 o 4 series no se traduce en una interpretación concisa por parte del usuario, sino que crea un exceso de puntos y colores, y tiende a crear confusión.

Una opción para solventar esto es representar cada una de las múltiples series de datos en un gráfico aparte de dimensiones reducidas y siempre jugando con dos o tres colores complementarios entre sí para que, al mostrarlos en paralelo, se consiga una mejor apreciación de las tendencias de cada una de las series de datos y se diferencien las dos o tres métricas.

### 3.5.3 Gráficos de caja

En tercer lugar, los *diagramas de caja* son muy convenientes para visualizar la distribución de datos con respecto a cuatro partes. Por un lado, las tres líneas que comprenden la caja son, en orden: primer cuartil, segundo cuartil (o mediana) y tercer cuartil. Por otro lado, las líneas extremo de la caja (bigotes) son el pico mínimo y el pico máximo de la distribución; y algunos puntos aislados permiten hacer comparaciones con los valores críticos de las series de datos.

Esto se plasma en la figura 22:

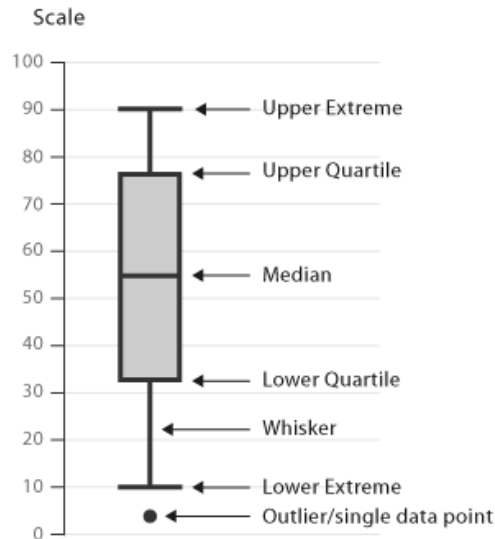


Figura 22. Anatomía de un gráfico de caja.

Cabe decir que este tipo de gráficos son supletorios de otras formas – como por ejemplo rectángulos o circunferencias – que representan los elementos pertenecientes a la distribución, tratándose las cajas como forma de análisis estadístico. Es por ello que, al complementarse con otras formas, el diagrama de cajas se caracteriza por tener colores monocromáticos y normalmente situados en la escala de grises, para así posicionarse en el frente del gráfico y a su vez no interrumpir a los otros.

Así pues, este tipo de gráficos optimizan la visualización de simetrías dentro de una serie de datos y, si se comparan múltiples series, se pueden contrastar las distribuciones de datos y los puntos críticos dentro de una población. Es por ello que pueden ser utilizados para mostrar desviación y así ser una alternativa muy aceptable a los gráficos de barras.

Por poner un ejemplo, este tipo de gráficos sería interesante en un elemento cuantitativo como bien es el nivel de sodio en la sangre. Mediante un diagrama de caja y bigotes sería posible contemplar la distribución de estos niveles y se distribuyen en el sistema sanguíneo de un organismo, contemplando los valores normales de la distribución, si está sesgado, si existen valores atípicos, y otros conceptos estadísticos personalizados.

### 3.5.4 Gráficos de dispersión

Por otro lado, los *gráficos de dispersión* son utilizados para determinar la correlación entre dos variables. Una de las variables se posiciona en el eje X mientras que la otra se posiciona en el eje Y, quedando los puntos en las coordenadas correspondientes del espacio cartesiano de acuerdo a dichas variables.

El aspecto puede ser, por ejemplo, según se muestra en la figura 23, donde la recta de regresión lineal establece el ajuste de la relación entre variables.

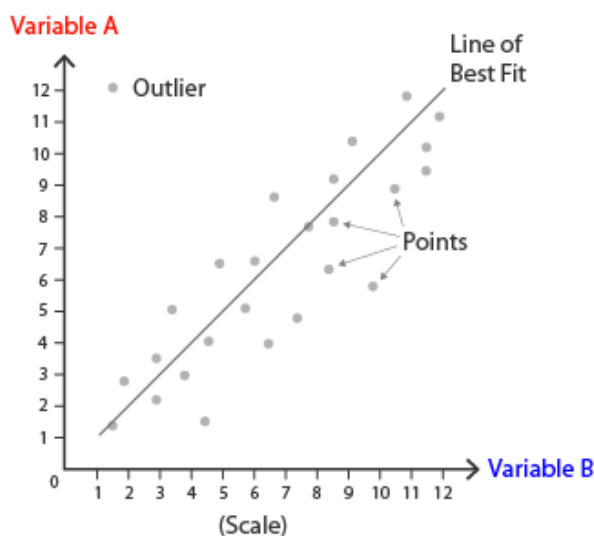


Figura 23. Anatomía de un gráfico de dispersión.

Cabe destacar que para el uso de este tipo de gráficos se ha de tener claro que las variables tomadas no son influidas por otras. Si estas dos variables se pueden parear de forma aislada es posible obtener un buen nivel de información de la relación de las variables en dos niveles:

1. La forma en la que la dispersión crea correlación entre las variables, que realmente puede ser lineal (positiva o negativa), no lineal, o ni siquiera existir. En caso de existir, la forma de la dispersión se traduce en la tendencia de la población con respecto a dichas variables, lo cual resulta de utilidad para crear aproximaciones mediante regresión.
2. El nivel de correlación entre los propios datos, de modo que, si los puntos dispersos se encuentran próximos entre sí, existe una mayor correlación que si están separados en el espacio. Es por ello que no existirá correlación cuando los puntos se encuentren totalmente dispersos por el espacio cartesiano.

Así pues, un ejemplo que podría abarcar este tipo de gráficos y que además resulta muy común es la búsqueda de relación entre peso y altura. Si se traza en el eje Y el peso y en el eje X la altura, se podría comprobar cada cuántas unidades de altura aumentan las unidades de peso y determinar por tanto el nivel de correlación entre estas dos variables para un segmento de población en concreto, así como los casos aislados que existen en la misma.

### 3.5.5 Gráficos de burbujas

Un tipo de gráficos funcionalmente similar al anterior es el *gráfico de burbujas*, aunque con dos notables diferencias. Por un lado, los puntos aquí se convierten en burbujas y éstas se categorizan por colores que siguen una regla de analogía (mostrados en una leyenda) para así tener una clasificación de las series de datos que realmente se quieren comparar en base a las variables. Por otro lado, el diagrama de burbujas permite añadir una variable extra al espacio de coordenadas cartesiano, la cual viene determinada por el área de la burbuja y establece una comparación entre elementos de la misma categoría.

Esto se puede contemplar en la figura 24:



Figura 24. Anatomía de un gráfico de burbujas.

Un ejemplo de utilización en este tipo de gráficos podría ser la búsqueda de comparaciones entre distintas dimensiones dentro de varios pacientes. Esto es, estableciendo en los ejes vertical y horizontal dos métricas, se podría jugar con atributos dentro de una misma población y asociarlos a otras métricas y dimensiones. Sin ir más lejos, comprobar cómo el peso y la altura (ejes vertical y horizontal respectivamente) juegan un determinado papel en un segmento de población donde existen personas intolerantes a la lactosa (caracterizado por el color de las burbujas) y con un determinado nivel de glucosa (caracterizado por el tamaño de las burbujas).

Así pues, un diagrama de burbujas correctamente representado permite incluso establecer patrones entre categorías y correlaciones entre las variables de una población; no obstante, es importante asegurarse de que no existan demasiadas burbujas para no saturar la interpretación del usuario. Una forma de paliar este problema es incluyendo en el gráfico la posibilidad de interacción por parte del usuario y selección de categorías a comparar.

### 3.5.6 Gráficos de pastel

Por otro lado, los *gráficos de pastel* son diagramas circulares muy útiles para hacer énfasis en las diferencias proporcionales entre categorías y así permitir al usuario visualizar las relaciones entre éstas y la cantidad total de datos.

Para ello, las categorías se caracterizan por colores y la proporción de datos con respecto al total es un trozo de “pastel” cuya área viene determinada por el arco radial. Esto quiere decir que la cantidad de una categoría con respecto al total se puede mostrar

como un porcentaje dentro del gráfico, y la dimensión de esta proporción se obtiene convirtiéndola a grados y teniendo en cuenta que la totalidad del pastel son 360 grados.

Un ejemplo de este gráfico se muestra en la siguiente figura:

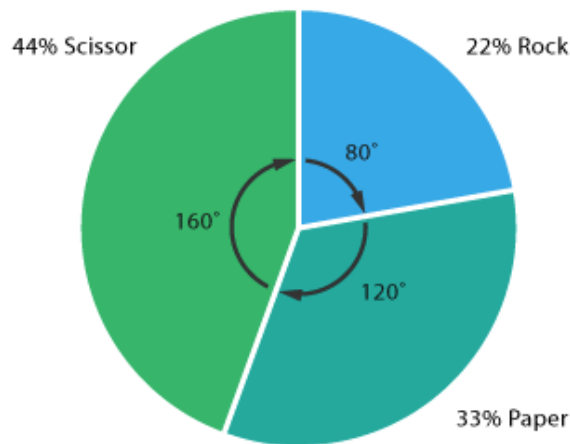


Figura 25. Anatomía de un gráfico de pastel.

Como principal punto débil, este tipo de gráfico no resulta conveniente para mostrar demasiadas categorías ya que el tamaño de los “trozos del pastel” se ven reducidos en exceso. Además, hay que tener en cuenta que se tratan de forma aislada, ya que una vez comparar un gráfico de pastel con otro no tiene demasiado sentido, pues visualmente se pierden las proporciones.

Ejemplos de utilización estarían enfocados a: distinguir el género de una determinada población, jugar con los rangos de edad (sin excederse en número), clasificar según tres o cuatro niveles el estado de una enfermedad, etc...

Es por ello que los gráficos de barras convencionales y de área apilada pueden resultar más útiles para comparar categorías, y probablemente optimicen más el espacio de muestra de los datos.

### 3.5.7 Diagrama de Venn

Un tipo de gráfico capaz de representar las relaciones lógicas entre conjuntos de datos son los *diagramas de Venn*. Cada conjunto de datos representa una categoría distinta, y estas categorías se plasman mediante circunferencias que separan sus datos en dos partes: por un lado, la parte que corresponde con la intersección del resto de categorías; y por otro, la parte que por el contrario no corresponde con otra categoría.

Esto se convierte en una relación de datos entre categorías, tal y como se muestra en la figura 26:

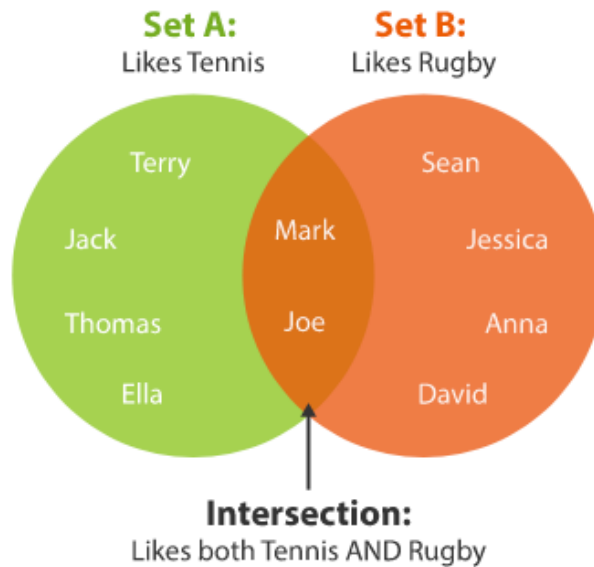


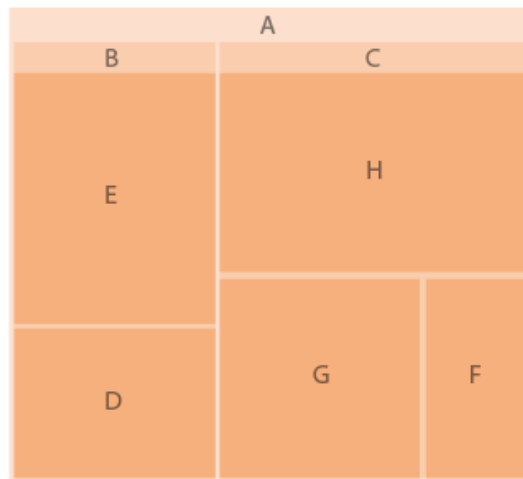
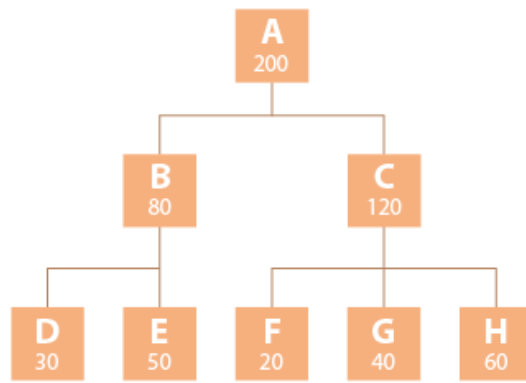
Figura 26. Anatomía de un gráfico de Venn.

Así pues, el área de intersección entre las circunferencias es el área de solape que indica un grupo de características en común entre categorías. Generalmente el número de categorías suele rondar entre 2 y 7, ya que a partir de ese número podría generarse una geometría demasiado compleja de interpretar. También hay que tener muy en cuenta el uso de colores en este tipo de gráficos, pues la zona de solape debe quedar totalmente diferenciada del exterior de las circunferencias.

Un ejemplo de aplicación del diagrama de Venn se podría encontrar en la cantidad de pacientes que sufren una enfermedad (determinados por una circunferencia) y a la vez sufren otra.

### 3.5.8 Mapa de árbol

En otro lugar, los *mapas de árbol* resultan una forma de categorizar la información de forma proporcional mediante áreas rectangulares y además crear subcategorías de forma que se puedan mostrar jerárquicamente las dimensiones del conjunto de datos en cuestión.



**Figura 27. Anatomía de un diagrama de árbol.**

Tal y como se observa, el tamaño de área de cada uno de estos rectángulos varía en función de la cantidad de datos que pertenezcan a su categoría correspondiente, por lo cual representa de manera perfecta la proporcionalidad de estas categorías. Otro aspecto muy importante es el color, que puede caracterizar otro tipo de dimensión siempre y cuando no se exceda demasiado, pues la heterogeneidad de dicha dimensión en la jerarquía podría causar confusión en los rectángulos, por lo cual es recomendable utilizar colores que cumplan la regla de analogía.

Además, también se puede ver que la categoría A se puede dividir según B y C, que a su vez tienen asociadas más subcategorías. El tamaño del área de la categoría padre es el total de subcategorías, de modo que si no se asocia cantidad a una subcategoría el área se divide de forma equitativa entre las otras.

Por ejemplo, para investigar las causas de una enfermedad en un determinado segmento de pacientes se podría partir de dos tipos de síntomas y a raíz de esto ir descendiendo el nivel de jerarquía hasta que se concrete el nivel punto de origen de las causas. Aquí se jugaría con el tamaño de los bloques en función del número de pacientes y existiría la posibilidad de añadir mediante el color una dimensión auxiliar que permitiera analizar estas causas de enfermedad.

### 3.5.9 Mapa geográfico

Por último, un tipo de gráfico particular que cabe destacar es el *mapa geográfico*. Este tipo de visualización es apropiada para dividir variables según regiones, existiendo una alta flexibilidad en la disposición de estas variables.

Por un lado, los datos se pueden mostrar coloreando las regiones geográficas para compararlas en función de un determinado rango de valores de una misma variable. Los colores que se suelen utilizar para esto son paletas correspondientes a uno o dos tonos jugando con la saturación y luminancia, para así tener un espectro visualmente reconocible desde el punto de vista cuantitativo.

Se representa este caso sobre la figura 28:

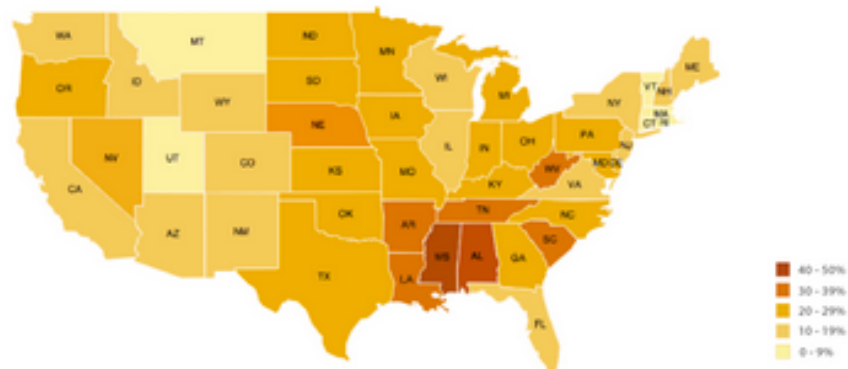


Figura 28. Anatomía de un ejemplo de mapa geográfico.

Un punto crítico en el tipo de mapa geográfico mostrado en la figura 28 es que el espectador tiende a relacionar de forma preconsciente las regiones más extensas con un valor o rango de valores más elevado que una región más reducida. Es por ello que los valores destacables en regiones pequeñas tienen dificultad de ser correctamente interpretados.

Una forma de paliar esto en un número reducido de regiones es mediante etiquetas numéricas que muestren las cantidades de la variable a representar. Estas etiquetas se engloban en burbujas para contemplar las proporciones entre cada una de las regiones. El hecho de verse limitado a una cantidad reducida de regiones se debe al posible solape entre burbujas.

Se muestra este otro caso en la figura 29:

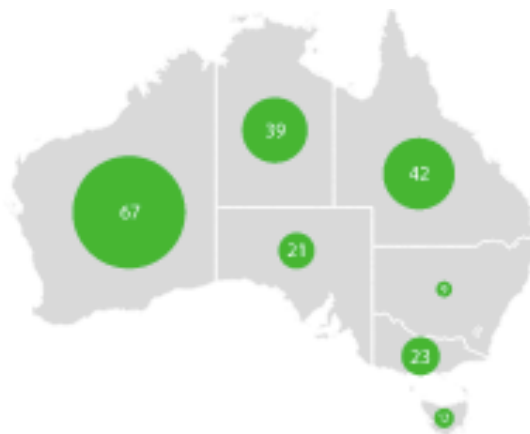


Figura 29. Anatomía de un ejemplo de mapa geográfico.

Así pues, mediante la anatomía de la figura 28 se podría plasmar el nivel de riesgo por región en el desarrollo de una enfermedad; mientras que gracias a la anatomía que se corresponde con la figura 29 se podría representar el número exacto de pacientes por región que ya posee una determinada enfermedad.

### **3.6 Recapitulación de la metodología genérica propuesta**

Mediante este capítulo se constata una metodología CBR genérica que puede ser implementada en cualquier dominio donde la tarea principal a resolver sea la recuperación de plantillas de informes; y a modo de sub-tarea dentro de este ciclo CBR general, se tenga como enfoque el poder reutilizar y adaptar elementos visuales acordes a los huecos de cada tipo de informe. Tal y como ya se ha dicho, este sub-ciclo CBR es precisamente la parte que abarca el proyecto.

Así pues, a pesar de haberse tratado en este capítulo la metodología de una forma totalmente genérica, en el siguiente se explica y analiza el experimento llevado a cabo en este proyecto y que supone la prueba de concepto de esta metodología. Este caso práctico permite particularizar el sub-ciclo CBR de reutilización y adaptación de elementos gráficos en el dominio de la Medicina del Deporte, concretamente en el entorno de análisis de pruebas EMG.

## Capítulo 4. Caso práctico de estudio y aplicación

La metodología comentada en el capítulo anterior podría tener como aplicación cualquier caso de generación de informes en el que se tenga como objetivo prioritario el hecho de reutilizar gráficos que combinen fuentes con datos personales y señales generadas. Mediante este cuarto capítulo se estudia el caso práctico tratado en el proyecto: la electromiografía como método de evaluación de los músculos cervicales y por consiguiente como fundamental técnica de trabajo entre los especialistas del campo de la Medicina del Deporte.

### 4.1 mDurance

En el transcurso de este proyecto se cuenta en todo momento con el soporte de una entidad, mDurance [48], cuyas líneas de trabajo se encuentran enfocadas a la salud digital. La herramienta propia de mDurance se dedica a la medición, evaluación y verificación del estado de los músculos mediante EMG para orientar los resultados a los profesionales de la salud física, el deporte y el bienestar en general. Ofrece un sensor portátil de electromiografía, una aplicación móvil para Android, almacenamiento Cloud y análisis de datos a través de una aplicación web.

El esquema de funcionamiento del sistema se puede contemplar a continuación:

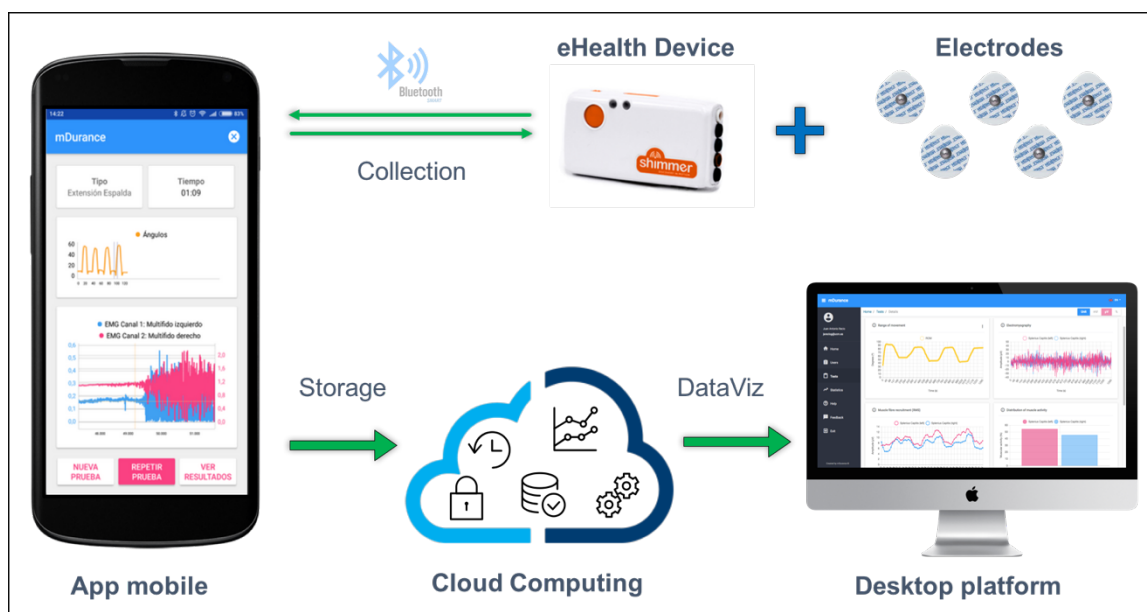


Figura 30. Esquema de funcionamiento del entorno de salud digital mDurance.

Como se puede observar en la figura 30, mDurance basa su mecanismo en la adquisición de la señal electromiográfica mediante un dispositivo colocado en el deportista/paciente que actúa como sensor a través de cinco electrodos. Seguidamente, estas señales se procesan y se filtran de ruido para almacenar los datos en el Cloud, además de proporcionar contenido gráfico en la plataforma para la visualización de dichos datos.

El poder real de estos datos consiste en que el experto pueda comprenderlos de forma precisa e interpretar los resultados para verificar anomalías musculares, evaluar el rendimiento muscular y detectar asimetrías musculares, entre otras muchas utilidades.

## 4.2 Exploración del contenido de la plataforma as-is

Llegados a este punto es importante que, antes de procesar, analizar y/o representar los datos tomados a los pacientes en nuestro estudio, es conveniente dar un repaso a la plataforma inicial de mDurance para explorar las funcionalidades y la forma que tiene de ofrecer los datos, y contrastar también con el abanico de necesidades que pueden surgir a un experto médico tras la obtención de datos procedentes de una prueba EMG.

### 4.2.1 Exploración de los modelos de informes iniciales

Hasta el momento mDurance trabaja con algunos prototipos de plantillas de informes mediante las cuales se ayuda a los profesionales médicos en la toma de decisiones, tanto en el diagnóstico como en el tratamiento de una posible patología. Es por ello que las entradas de datos que se sugieren en dicha plantilla deben proporcionar la información completa de la prueba realizada al paciente, tanto desde el punto de vista sintáctico como semántico.

Los prototipos de informes se muestran en la figura 31, y las distintas partes en las que se desglosan pueden resumirse en:

- Tipo de informe mediante el cual se describe la prueba muscular, el paciente en cuestión, la fecha y la caracterización del usuario al cual se dirige (como bien puede ser un doctor, el propio paciente o un científico de datos).
- Las principales funciones del informe, el estado actual del paciente, su evolución, así como una comparación con el contexto para cubrirse en caso de detección de fraude.

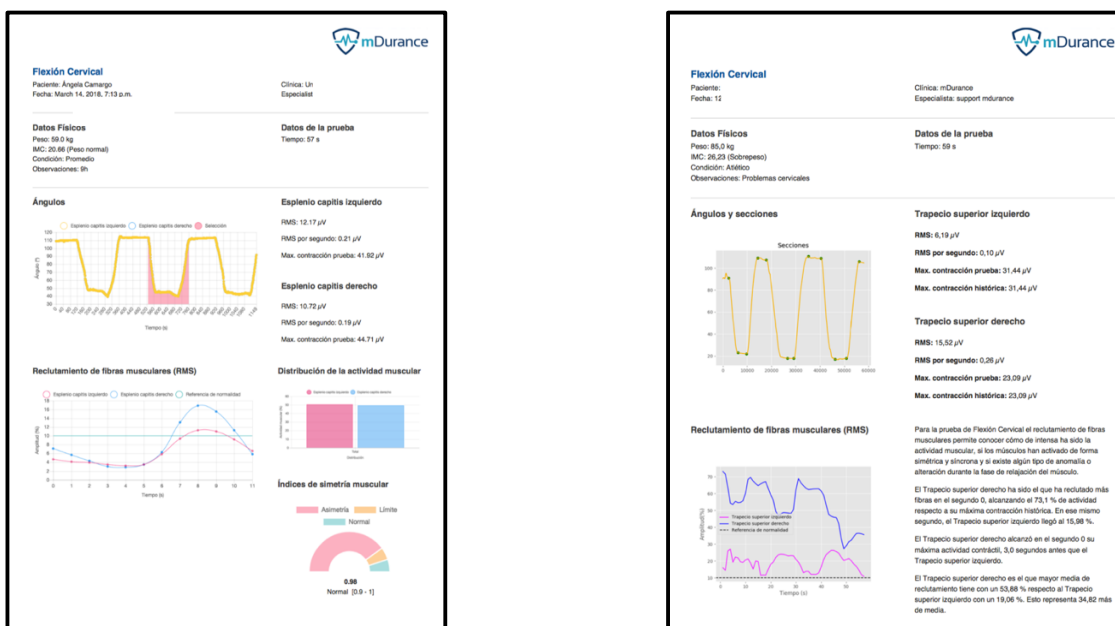


Figura 31. Prototipos de informes utilizados inicialmente en la plataforma de mDurance.

Desde un punto de vista funcional, se puede comprobar que estos informes iniciales explican correctamente los datos y permiten a un usuario deportista no experto en el campo de la medicina el hecho de poder verificar anomalías musculares, evaluar el rendimiento de los músculos y detectar asimetrías en ellos.

No obstante, estas explicaciones visuales son generadas de forma aislada; es decir, no se encuentran asociadas a un contexto en el que se muestren unos patrones personalizados, ni tampoco plasma una evolución temporal, ni mucho menos la explicación del motivo por el cual el paciente en cuestión posee (o no) una patología igual (o distinta) al resto de pacientes que poseen unas características contextuales parecidas.

Es por esto que existe un área de mejora muy interesante a la hora de presentar y comunicar los datos de forma efectiva en nuestro caso de estudio de mDurance. Los datos sintácticos de los pacientes interpretados a modo de señales EMG se pueden relacionar con la información semántica (o contextual) del mismo u otro paciente, así como con otras características poblacionales.

Esto se traduce en un grado de personalización más elevado ya que el consumidor del informe puede poseer distintos puntos de comparación y comprender así cómo los datos se relacionan bajo determinados factores dentro del contexto de la electromiografía.

#### **4.2.2 Exploración de contenido de datos**

La plataforma de mDurance expone sus datos en función de secciones bien separadas. Se describen a continuación.

Por un lado, el apartado de **estadísticas** es el más general y en él se muestra el número de pruebas clasificadas según ciertos aspectos comunes entre la población de pacientes, como bien son la edad, el sexo y el estado muscular; mostrado la totalidad y el porcentaje de la suma de pruebas.

Por otro lado, El apartado de **usuarios** almacena algunos datos básicos acompañado de un pequeño resumen donde se dispone de las pruebas realizadas para cada uno de los pacientes.

- Datos físicos del paciente, como su IMC, peso, condición física y observaciones.
- Registro de valores para cada una de las pruebas realizadas, con la opción de filtrar por músculo y unidad (microvolts, milivots y %). Entre los parámetros se encuentran los siguientes:
  - Tiempos de duración.
  - Promedio del nivel de actividad muscular (RMS).
  - Promedio del nivel de actividad muscular por segundo (RMS/s).
  - Máximo nivel de contracción muscular (MVC).

Por último, se encuentra el apartado que unifica toda la base de datos de las **pruebas** realizadas por cada uno de los pacientes. Aquí se listan cada una con información sobre la fecha de realización, el paciente, el tipo de prueba y todos los parámetros tomados en el experimento. Estos parámetros no solamente definen en mayor profundidad la condición muscular del paciente, sino que también poseen los niveles neuromusculares de la prueba EMG. Esto se resume en:

- Datos físicos del paciente, como su IMC, peso, condición física y observaciones.

- Duración de la prueba y detalles de los dos músculos evaluados:
  - Nivel de dolor.
  - Nivel de fatiga muscular.
  - Sufrimiento de lesión.
- Registro de la actividad muscular llevada a cabo en la prueba EMG, donde también se permite filtrar por unidad (microvolts, milivots y %). Se divide en los siguientes puntos:
  - Rango de movimientos, donde se muestran la continuidad de los ángulos asociados a la actividad muscular en ventanas temporales de 0,05 segundos.
  - Distribución de los niveles de tensión muscular en ventanas temporales de 0,05 segundos.
  - Reclutamiento de fibras musculares, donde se muestra la distribución de los niveles de tensión en ventanas temporales de 0,91 segundos.
- Distribución de la actividad muscular que ha ejercido cada uno de los músculos en la ejecución de la prueba, para así determinar también el grado de asimetría de éstos.
- Por último, se recogen algunos valores promedios para cada uno de los dos músculos evaluados en el experimento. Entre ellos se encuentran el RMS, RMS/s, así como el máximo contracción de esta prueba y del histórico de pruebas.

A modo de síntesis, en esta exploración se observan una serie de carencias que hacen que mDurance no explote toda su funcionalidad en este aspecto. La principal causa radica en que la cantidad de etiquetas de contexto no es lo suficientemente amplia, ya que el máximo nivel de abstracción que se tiene es el dolor del paciente y su condición física. No se inspeccionan las causas de estos dolores mediante la comparación con otras etiquetas contextuales, ni tampoco se buscan patrones de correlación con datos sintácticos.

Por tanto, constatamos que el modelo de caracterización de datos propuesto en el presente proyecto y descrito en el apartado 4.6 no solamente incluye estas etiquetas de datos, sino que también se encarga de complementarlas con otras etiquetas y mediante la representación gráfica buscar patrones que expliquen la relación entre características sintácticas de la manera más eficiente posible.

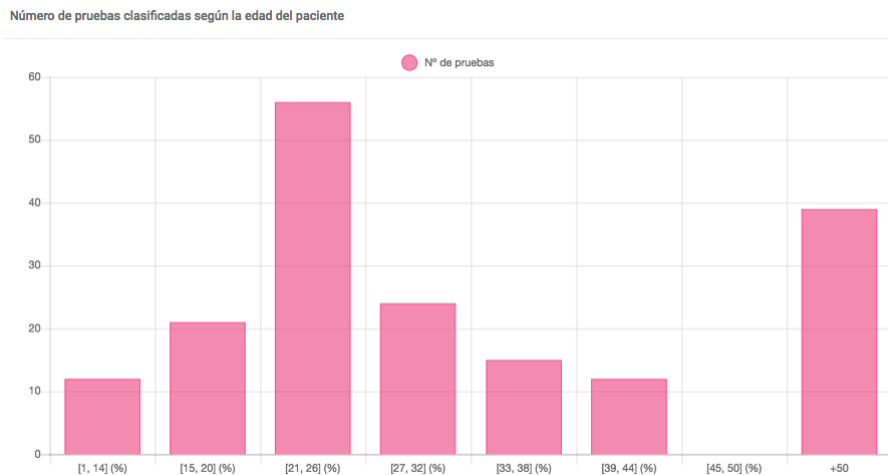
### **4.2.3 Exploración de elementos visuales**

De forma general, mDurance expone los datos anteriormente mencionados a través de elementos visuales, viéndose suplementados con etiquetas de texto y breves descripciones sobre los valores que resultan más comunes.

Desde el punto de vista del diseño gráfico, el estado actual de la plataforma digital dispone de múltiples puntos que son susceptibles de mejora. El punto más destacable es la escasa variedad de gráficos utilizados. En realidad, únicamente son implementados gráficos genéricos *de líneas*, para la representación de señales en el tiempo; *de barras*, para cuantificar datos y compararlos; y *de pastel*, para mostrar proporciones.

A continuación, se comentan algunos ejemplos de elementos visuales donde se detectan otros puntos débiles importantes en esta generación de gráficos.

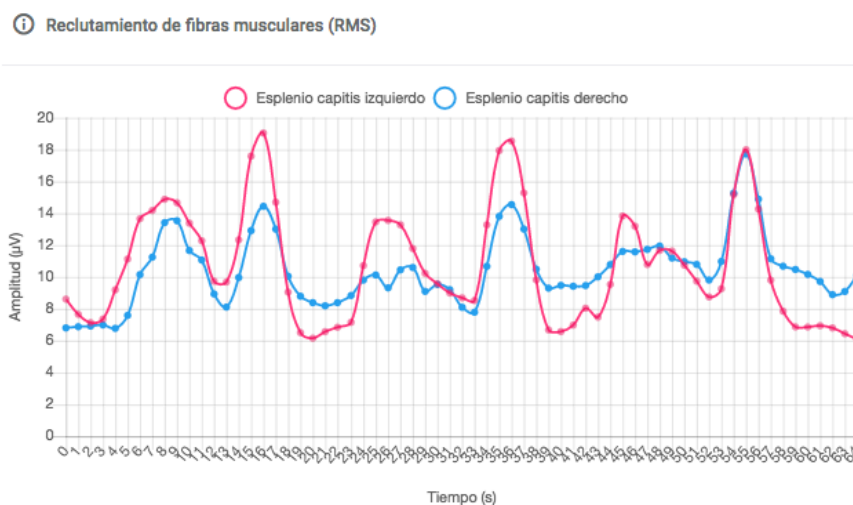
En primer lugar, no se estructuran los gráficos de forma que se aprovechen en su totalidad, sino que ciertas capacidades funcionales se dejan “vacías”. Sin ir más lejos, en la siguiente figura se puede observar que se muestra un solo color que determina el número de pacientes. Esto en realidad no tiene mucho sentido ya que el número de pacientes debería de ser cuantificable, y un color no lo es. El número de pacientes se muestra en el eje vertical, mientras que el atributo del color es una capacidad del gráfico muy aprovechable para mostrar otras características de los pacientes.



**Figura 32. Representación as-is del número de pruebas por edad del paciente.**

En segundo lugar, existen otros detalles en la representación de los ejes como por ejemplo la acumulación de números en cada una de las líneas temporales del eje horizontal, lo que causa al visualizador incapacidad para detectar el patrón de la serie temporal. Además, imposibilita el hecho de poder hacerse una idea del transcurso de los movimientos FRP de forma preconsciente, ya que es necesario indagar en cada uno de los puntos del eje horizontal para poder asimilar los instantes de tiempo.

Se muestra a través de la figura 33:



**Figura 33. Representación as-is del RMS en una prueba de paciente.**

En tercer lugar, es importante destacar también la existencia de un abanico de colores altamente limitado. En los gráficos de barras, se juega prácticamente en todo momento con el color magenta, lo cual supone una visualización demasiado monótona para el espectador.

Esto se puede apreciar en la figura siguiente:

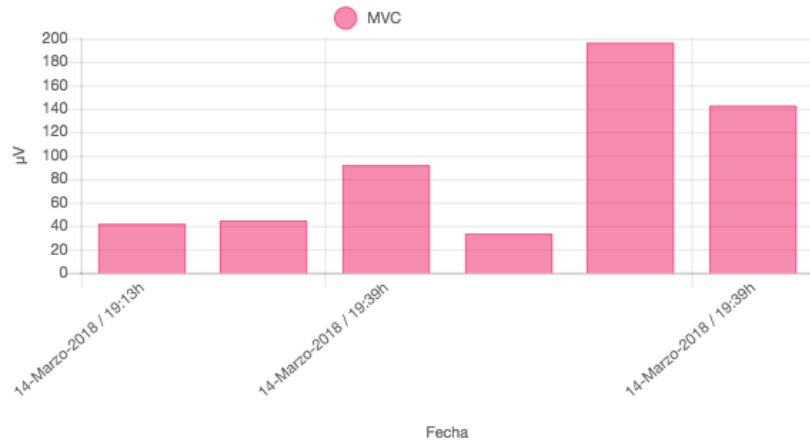


Figura 34. Representación as-is del MVC por cada prueba que un paciente realiza.

Por último, también se puede destacar la ausencia de una paleta de colores determinada. En su lugar, los colores se mezclan sin seguir un patrón claro, se repiten tonos de colores, y no se aprovecha en ningún momento de dos componentes clave dentro de la teoría del color como son la saturación y la luminancia.

Esto provoca que en ocasiones donde se tienen demasiados valores, se (se mezclan sin sentido, se repiten colores) lo cual supone que la visualización en ciertas ocasiones sea algo monótona e incluso tediosa.

Lo vemos en el gráfico de pastel de la figura 35:



Figura 35. Representación as-is del número de pruebas según el estado físico del paciente.

### 4.3 Experimentación

El propósito principal a partir de este punto comprende el estudio y potencialización de los componentes visuales de la herramienta de salud digital mDurance. Para ello, el proyecto en sí se basa en un experimento mediante el cual en primera instancia se pretende explotar la funcionalidad *as-is* de dicha herramienta para a posteriori potenciar su evolución *to-be* desde la perspectiva DataViz.

En el presente experimento se sigue la metodología descrita en el capítulo 3, mediante la cual se propone generar explicaciones visuales personalizadas en plantillas de informes de manera automática y con la opción de aprendizaje a través del comportamiento interactivo del profesional de la medicina del deporte.

Previamente al análisis de las formas de visualización llevadas a cabo en el proyecto, conviene justificar este contexto, la muestra de población que participa en el estudio, así como el material utilizado y las pruebas realizadas a cada uno de los pacientes.

#### 4.3.1 Características del estudio

El presente experimento tiene como punto de origen un problema de salud muy común en la población y que lleva de la mano múltiples líneas de investigación. Éste no es otro que el dolor cervical [49], y es de tal importancia que supone un factor de riesgo desde el punto de vista físico, psicosocial e individual, categorizándose en la sociedad mediante factores de todo tipo y también hábitos donde se precisen posturas estáticas prolongadas y/o intensas.

De hecho, algunos de los estudios más comunes llevados a cabo años atrás han demostrado que los resultados en este tipo de ejercicios cervicales no son los mismos en muestras de pacientes donde varía la edad, el género [50], el estrés rutinario, la ocupación [51], o incluso determinados hábitos de vida como el frecuente transporte de mochila [52]. Esto justifica que tanto investigadores como expertos médicos persigan constantemente la exploración de patrones estadísticos mediante pruebas EMG con tal de poder asociarlos a características comunes entre muestras de pacientes tomadas en un mismo contexto.

#### 4.3.2 Características de la población

El estudio evalúa el estado de los músculos cervicales en una población de 32 sujetos de entre 13 y 77 años, 12 mujeres y 20 hombres, todos ellos con características físicas y hábitos de vida lo más diversos entre sí para contrastar distintos factores en una misma población. Esta heterogeneidad de características es un factor indispensable en el experimento ya que en él se persigue, además de las etiquetas sintácticas obtenidas mediante el análisis de señales EMG, un correcto etiquetado semántico para así poder dotar al contexto de valor calificable.

Una forma de conseguir esto es incluyendo en el estudio algunos casos con patologías neuromusculares previamente detectadas en algunos de los pacientes, así como ciertas alteraciones entre sus costumbres diarias que pudieran suponer un punto crítico en la actividad cervical.

En el siguiente esquema se define la relación del campo sintáctico recogido en las señales EMG con las características semánticas de la población en el contexto médico:

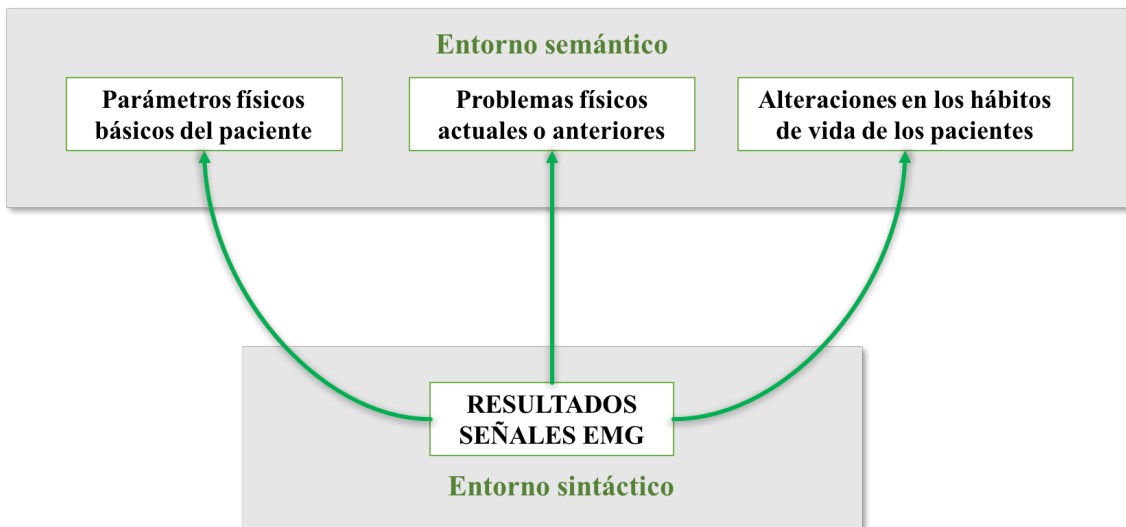


Figura 36. Relación entre el entorno sintáctico y semántico aplicable a la población del estudio.

### 4.3.3 Instrumentación y tecnologías

El material utilizado para la realización de este experimento se puede dividir en dos partes. Por un lado, el dispositivo IoT utilizado para la recolecta de los datos; y por otro, las plataformas tecnológicas encargadas de la ingesta, tratamiento y visualización de los datos.

Visto este desglose se comentan ambas partes a continuación.

#### 4.3.3.1 Dispositivo IoT utilizado

El dispositivo de recolecta de datos utilizado en el experimento es proporcionado por mDurance y supone el punto de interacción con cada uno de los pacientes. Consiste en un dispositivo electromiográfico Shimmer3 EMG [53] de pequeñas dimensiones y peso, pero con una alta capacidad de autonomía.

Este dispositivo posee unas características técnicas que hacen que cuente con 10 grados de libertad (DOF) vía acelerómetro, giroscopio, magnetómetro y altímetro, cada uno de ellos con un amplio rango de datos seleccionable que permite la lectura de datos con un nivel de ruido mínimo.

El funcionamiento de este wearable se basa en la recolección de los impulsos eléctricos musculares con un sistema A/D en señales de 16 bits, frecuencia de muestreo de 24 MHz y ancho de banda de 8,4 kHz. El dispositivo ofrece cinco inputs para conectar los electrodos encargados de capturar el test EMG:

- Dos canales laterales, izquierdo y derecho, donde se permiten conectar cuatro electrodos (dos por canal) y medir los niveles de señal de forma simétrica en los músculos cervicales.
- Un electrodo situado en el centro del dispositivo y correspondiente a la toma de tierra.



**Figura 37. Conexiones del wearable Shimmer3 EMG.**

Las especificaciones técnicas del wearable Shimmer3 EMG quedan adjuntas en el Anexo 1 de este documento.

#### **4.3.3.2 Tecnologías empleadas**

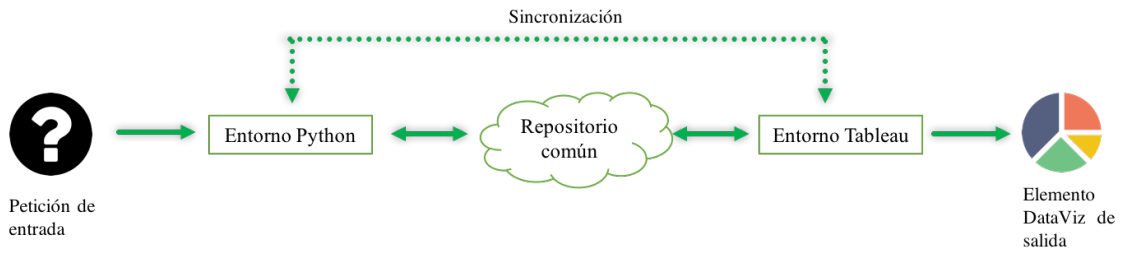
Claro está que desde que el dispositivo Shimmer3 EMG recoge los datos de un paciente hasta que un tipo de gráficos es reutilizado y adaptado, todo el conjunto de datos lleva a cabo una serie de procesos que no resultan ni mucho menos triviales. No es objeto de este estudio contemplar una comparativa entre plataformas tecnológicas, ni por la parte de tratamiento ni tampoco por la de visualización de datos; no obstante, sí resulta importante realizar una breve síntesis que justifique el procedimiento seguido desde que se obtienen los conjuntos de datos hasta que se representan.

Cabe decir que las plataformas utilizadas para la realización de este proyecto se ejecutan de forma local a modo de “banco de pruebas” pero siempre teniendo en mente la migración Cloud del sistema en su totalidad.

Los entornos software implicados dentro de la arquitectura tecnológica son:

- Por un lado, Jupyter [54], desarrollado en Python y dedicado al manejo de los datos como plataforma maestra debido a la elevada flexibilidad que ofrece para el manejo de múltiples ficheros con arrays de dimensiones considerables.
- Por otro, Tableau [55], cuyo sistema de consultas se basa en SQL y se encarga de la posterior visualización de los datos de forma avanzada a partir de las fuentes de datos recogidas.
- Por último, un repositorio en común para la correcta sincronización de las dos plataformas, el cual facilita la comunicación de los datos y los archivos temporales con gráficos embebidos.

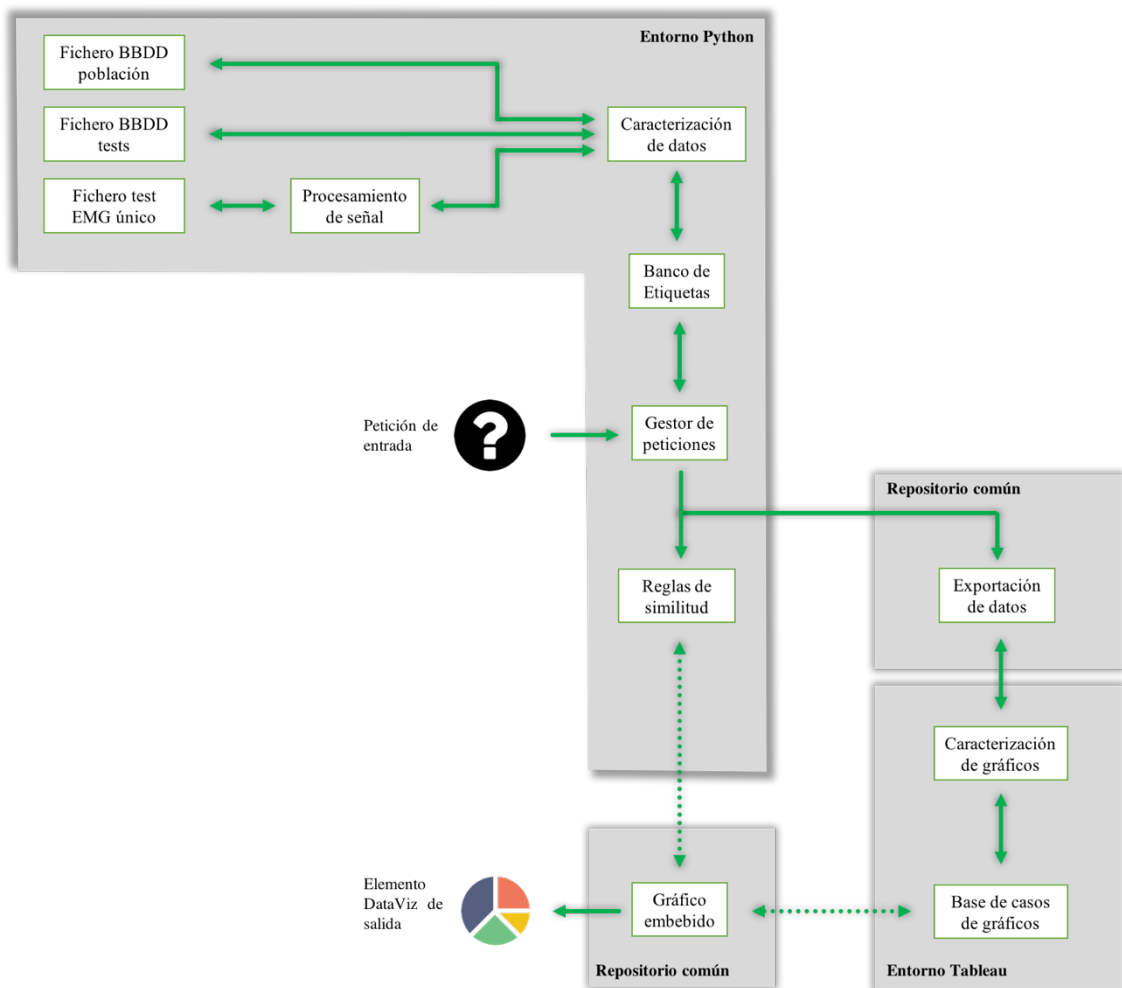
El flujo de datos dentro de la arquitectura de la solución propuesta se muestra de forma básica mediante el siguiente diagrama, correspondiéndose con los componentes software de la solución propuesta:



**Figura 38. Esquema básico de los flujos de E/S en el procedimiento seguido.**

En el anterior esquema se puede observar que el experto médico es capaz de realizar una petición de entrada al sistema, ésta ser atendida y gestionada por el entorno Python, y seguidamente comunicar al entorno Tableau el elemento gráfico a reutilizar en función de los datos transferidos, la petición del experto y las reglas de similitud establecidas.

La descripción detallada de los flujos E/S en la arquitectura software se desarrollan a continuación, distinguiendo el entorno Python, el repositorio común y el entorno Tableau:



**Figura 39. Diagrama detallado de los flujos de E/S en el procedimiento seguido.**

Por un lado, la secuencia de procesos que conforman el entorno Python queda resumida en:

1. Importación de datos desde el dispositivo Shimmer3 EMG e incorporación inmediata a la base de datos de pacientes y pruebas; así como almacenamiento de los ficheros individuales de cada una de las pruebas realizadas.
2. Procesado y limpiado los datos procedentes del análisis EMG para disponer de señales útiles, teniendo en cuenta en todo momento el punto de vista analítico contrastado desde diversas fuentes de investigación en el contexto.
3. Incorporación al modelo de caracterización de los datos para la creación de un banco de etiquetas capaz de satisfacer las peticiones de entrada.
4. Ejecución de reglas de mapeo y similitud de datos, así como exportación de los conjuntos al repositorio, que posee la función de nexo entre esta plataforma de manejo y la plataforma encargada de la visualización de los mismos.

Por otra parte, el entorno Tableau se caracteriza por ingerir los datos de forma continua y adaptarlos a las formas de visualización caracterizadas dentro de la base de casos que está almacenada en este entorno. Esto quiere decir que cualquier *pull* por parte del entorno Python va a ser procesado de forma inmediata por este entorno DataViz en función de las reglas de similitud que atienden la petición de entrada generada por el experto médico.

Es por esto que los gráficos se reutilizan y reajustan siempre que; por la parte de Python, los conjuntos estén correctamente caracterizados; y por la parte de Tableau, los gráficos resulten óptimos también en su proceso de caracterización.

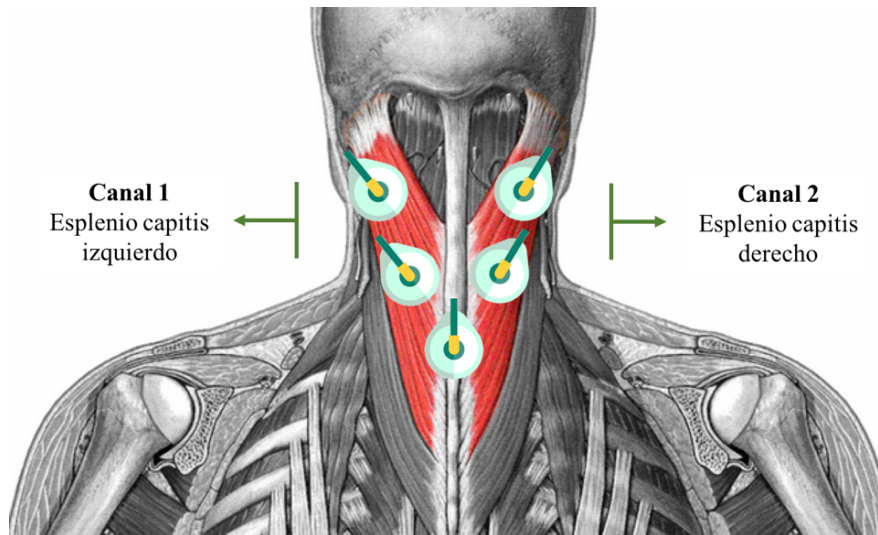
#### **4.3.4 Movimientos**

Una vez definida la muestra de población participe del experimento y el material con el que se lleva a cabo, se procede a explicar el funcionamiento del experimento como tal: los movimientos de los pacientes para cada una de las pruebas.

El rendimiento de la flexión que determina el dolor cervical está totalmente ligado al fenómeno de flexión-relajación (FRP), que es un método utilizado para evaluar las alteraciones en patrones de actividad muscular [56]. Este fenómeno se caracteriza por la ausencia de señales EMG por parte de los músculos en el movimiento de máxima flexión cervical, y en el presente estudio supone la tarea principal de cada uno de los pacientes en sus respectivas pruebas.

Para la correcta ejecución del experimento el dispositivo Shimmer3 se posiciona en la parte trasera de la cabeza del sujeto a través de una banda de goma elástica, y se colocan el canal 1 y 2 en los músculos cervicales, que respectivamente corresponden con los denominados *esplenio capitis izquierdo* y *esplenio capitis derecho*.

Esta disposición resulta de alta importancia a la hora de posteriormente contemplar la simetría de los músculos, por lo que se esquematiza a través de la figura 40:



**Figura 40.** Distribución de electrodos en función de los canales del Shimmer3 sobre los dos músculos.

Así pues, las pruebas FRP que miden la capacidad cervical de dos músculos se clasifican según movimientos isotónicos e isométricos. El experimento consta de tres tests para cada uno de los pacientes que se han tomado en el estudio, uno isotónico y dos con carga isométrica. Es por ello que este experimento consta de 177 pruebas en total.

Se describen a continuación los movimientos isométricos e isotónicos que los pacientes llevan a cabo en el estudio.

#### **4.3.4.1 Pruebas de movimientos isotónicos**

1. La primera y principal prueba se corresponde con un estiramiento FRP de tipo isotónico dividido en tres series. Mediante la figura 41 se ejemplifica una de las series:
  - **Fase a:** En primer lugar, la paciente se posiciona con la espalda totalmente recta y en una posición estática durante 5 segundos.
  - **Fase b:** Seguidamente, la paciente realiza una flexión total de las cervicales realizando un movimiento donde sitúa la barbilla sobre el pecho en un intervalo de 5 segundos aproximadamente.
  - **Fase c:** Después de esto, la paciente mantiene la cabeza en el punto de flexión máximo de las cervicales, otros 5 segundos.
  - **Fase d:** Para acabar la serie, la paciente realiza un movimiento de re- extensión hacia arriba en otro intervalo de 5 segundos hasta tener los músculos cervicales relajados y de esta forma la cabeza totalmente recta (vuelta a fase a).

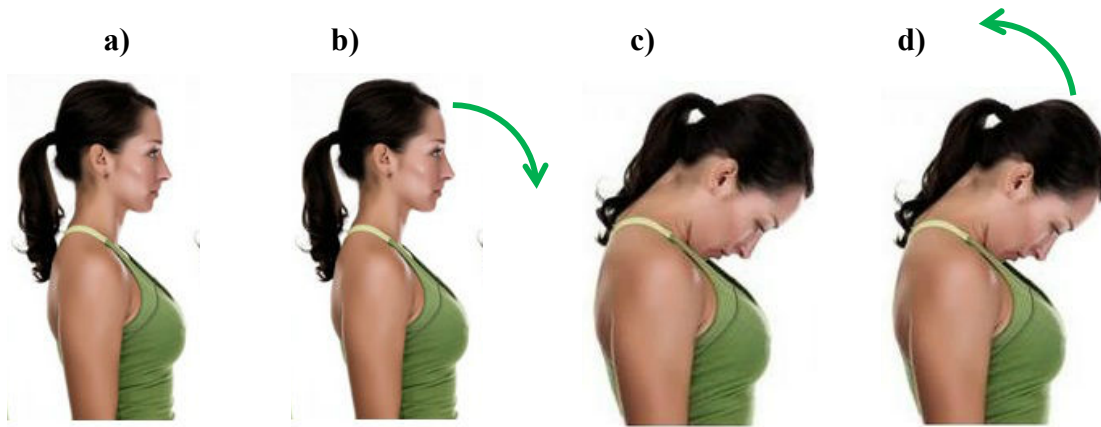


Figura 41. Secuencia de movimientos isotónicos FRP hacia delante y hacia atrás: Fases a-d.

Así pues, la duración de cada prueba de este tipo es:

$$((5 \text{ segundos} * 4 \text{ movimientos}) * 3 \text{ series}) + 5 \text{ segundos}_{final} = 65 \text{ segundos} \quad (4.1)$$

#### 4.3.4.2 Pruebas de movimientos isométricos

2. Por otro lado, el segundo test FRP se corresponde a un movimiento isométrico único de 10 segundos mediante el cual la paciente, con la cabeza recta, trata de flexionar las cervicales hacia adelante con la oposición de la mano (no de ella misma, sino del encargado de realización de la prueba). Esto se contempla en la figura 42 d.
3. Por último, de la misma forma que anteriormente, la paciente trata de flexionar la cabeza en un intervalo de 10 segundos, esta vez hacia detrás, con la resistencia de la mano en la parte trasera. Se observa en la figura 42 e.

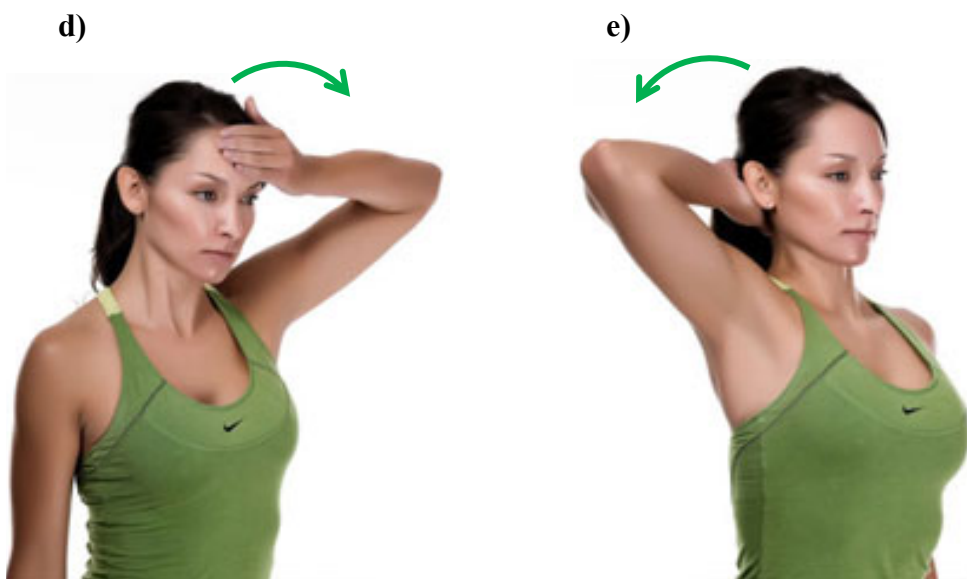


Figura 42. Secuencia de movimientos isométricos FRP hacia delante y hacia atrás.

#### 4.4 Procesamiento de datos

Independientemente del tipo de prueba ejecutada, cada una de ellas equivale a un fichero de datos en crudo, los cuales se toman como una secuencia de arrays. Entre estos destacan:

- La línea temporal determinada por la cantidad de milisegundos de la prueba.
- El número identificador del test.
- El tiempo total de la prueba, en segundos.
- El ángulo en el que se encuentra el Shimmer3 en cada instante de tiempo.
- Los niveles de señal en cada instante de tiempo, para los dos músculos sobre los que se realiza la prueba. Éstos se corresponden con los dos canales del Shimmer3 descritos anteriormente.

Para poder representar los valores sintácticos que proporcionan los niveles de señal EMG, se ha de conocer de primera mano la naturaleza de estas señales. Esto es, qué tipo de forma de onda se tiene originalmente y en qué se pretende convertir para tener conocimiento de los cálculos estadísticos y probabilísticos a realizar sobre las mismas.

A modo de breve ejemplo, se muestran en la siguiente figura un periodo de 20 ms de muestreo dentro de los 10 segundos que determina una prueba isotónica hacia atrás.

	id	total_time	name_muscle_1	name_muscle_2	angles	emg_muscle_1	emg_muscle_2
0	1714.0	10.0	Splenius Capitis (left)	Splenius Capitis (right)	75.5	1.932534	3.286539
1					75.49	1.938689	3.289616
2					75.49	1.938689	3.292693
3					75.48	1.941766	3.295771
4					75.48	1.947921	3.295771
5					75.47	1.954075	3.298848
6					75.47	1.957153	3.295771
7					75.46	1.972539	3.295771
8					75.46	1.981771	3.295771
9					75.46	1.987925	3.298848
10					75.45	2.000234	3.305002
11					75.45	2.003312	3.30808
12					75.44	2.000234	3.30808
13					75.43	2.006389	3.301925
14					75.43	2.006389	3.301925
15					75.43	2.006389	3.305002
16					75.42	2.015621	3.305002
17					75.42	2.024853	3.311157
18					75.41	2.024853	3.305002
19					75.41	2.034085	3.314234
20					75.4	2.034085	3.311157

Figura 43. Parte del conjunto de datos de una prueba EMG cualquiera.

Enfocando los datos desde la perspectiva que se encarga del análisis sintáctico, tenemos que los arrays que contienen los valores de cada músculo no son útiles tomados en crudo, ya que representan cualquier tipo de ruido procedente del entorno junto al nivel de señal eléctrica del músculo con el que el electrodo se encuentra en contacto.

En realidad, la naturaleza de las señales EMG no son fáciles de comprender, ya que son un cúmulo de variaciones en amplitud, frecuencia y fase; con una suma de ruido añadido. En nuestro caso, el nivel de ruido recogido por cada uno de los electrodos suele ser más o menos el mismo dentro de unos márgenes, pero el nivel eléctrico de cada músculo sí que es dependiente de la posición exacta en la que se encuentre el electrodo.

De cualquier forma, los niveles de ruido y variaciones en los niveles electromiográficos han de ser paliados de la forma más óptima posible con tal de poder visualizar los datos sintácticos correctamente. Qué mejor forma de comprobar esto que intentando representar los valores neuromusculares del anterior test en crudo:

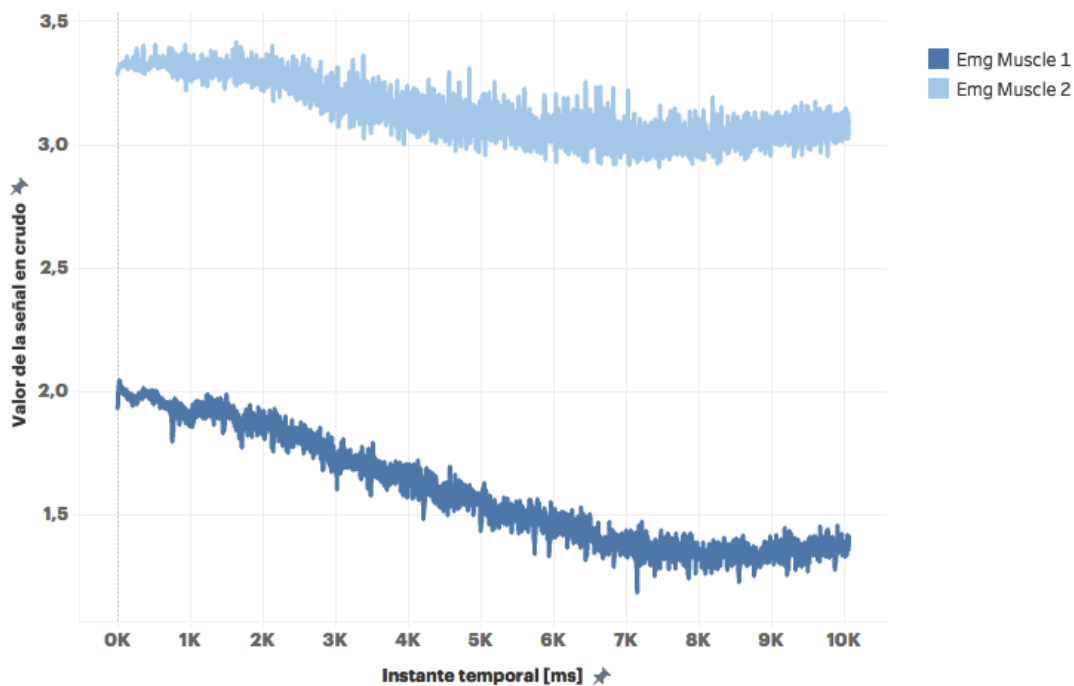


Figura 44. Señales en crudo correspondientes a los dos músculos cervicales medidos por el Shimmer3 en una prueba de 10 segundos.

Tal y como se puede comprobar, las visualizaciones de los valores que proporcionan los dos arrays EMG originales carecen de sentido si no son transformados en nivel útil de señal. Por tanto, se puede afirmar que el análisis sintáctico de estas señales prescinde absolutamente de las técnicas de procesado adecuadas.

Es por este motivo por el cual dichas técnicas se describen a continuación en líneas generales.

#### 4.4.1 Filtrado

En primer lugar, se aplica un filtro paso banda Butterworth con tal de producir la respuesta más plana posible hasta la frecuencia de muestreo de 1024 Hz. Dicho filtro tiene orden 4 y la banda pasante se genera desde los 20 Hz hasta los 450 Hz, de modo que se tenga una respuesta en frecuencia como la indicada en la figura:

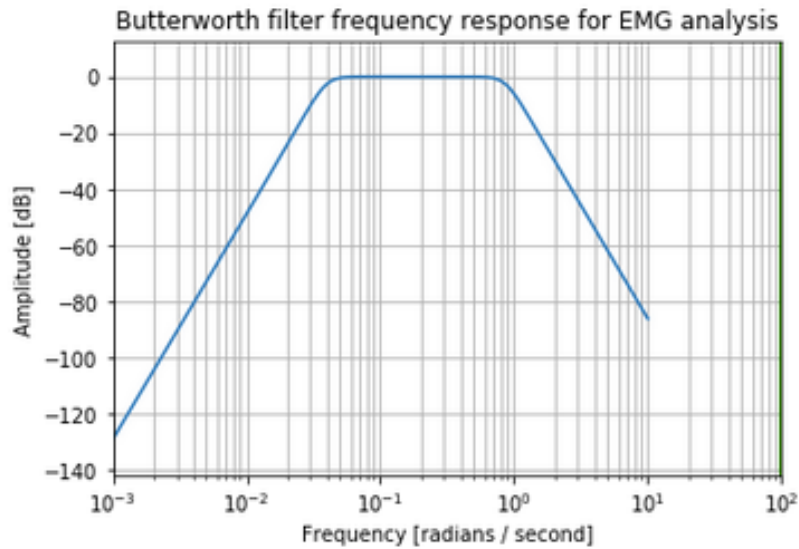


Figura 45. Respuesta en frecuencia del filtro Butterworth utilizado para el procesado de señales EMG.

De forma seguida, a la señal se le aplica un filtro de fase lineal en cada una de los extremos (determinados por las frecuencias de corte) preservando los picos de amplitud de lo que serían cada una de las bandas de frecuencia. Un ejemplo de respuesta de dicho filtro es el mostrado en la siguiente figura:

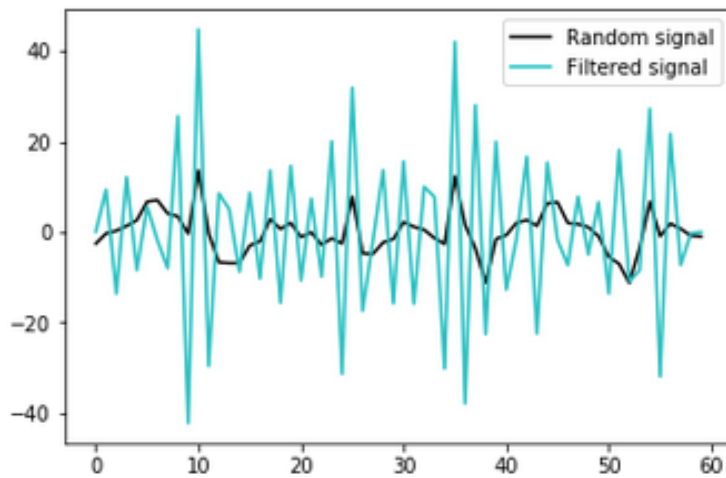


Figura 46. Ejemplo de filtro de fase lineal utilizado para el procesado de señales EMG.

Así pues, gracias a la utilización de estos dos filtros se consiguen eliminar las frecuencias altas y bajas a partir de las respectivas frecuencias de corte sin perder el formato de la señal EMG en cuestión. No obstante, esto no implica que dicha señal se encuentre libre de ruido.

#### 4.4.2 Eliminación de ruido

El ejercicio que se lleva a cabo en el último paso del procesado se resume de la siguiente forma: la señal EMG previamente filtrada se divide en tramos para calcular las desviaciones de cada uno de ellos y detectar los niveles de ruido. De este modo, la desviación de cada tramo se inventana como señal útil en caso de tener la suficiente amplitud, y se sustituye por tramos con amplitud de nivel 0 en caso de tratarse de tramos con ruido.

Con esto, las señales representadas en el ejemplo de la figura 44, antes en crudo, se muestran ahora tras el filtrado y eliminado de ruido:

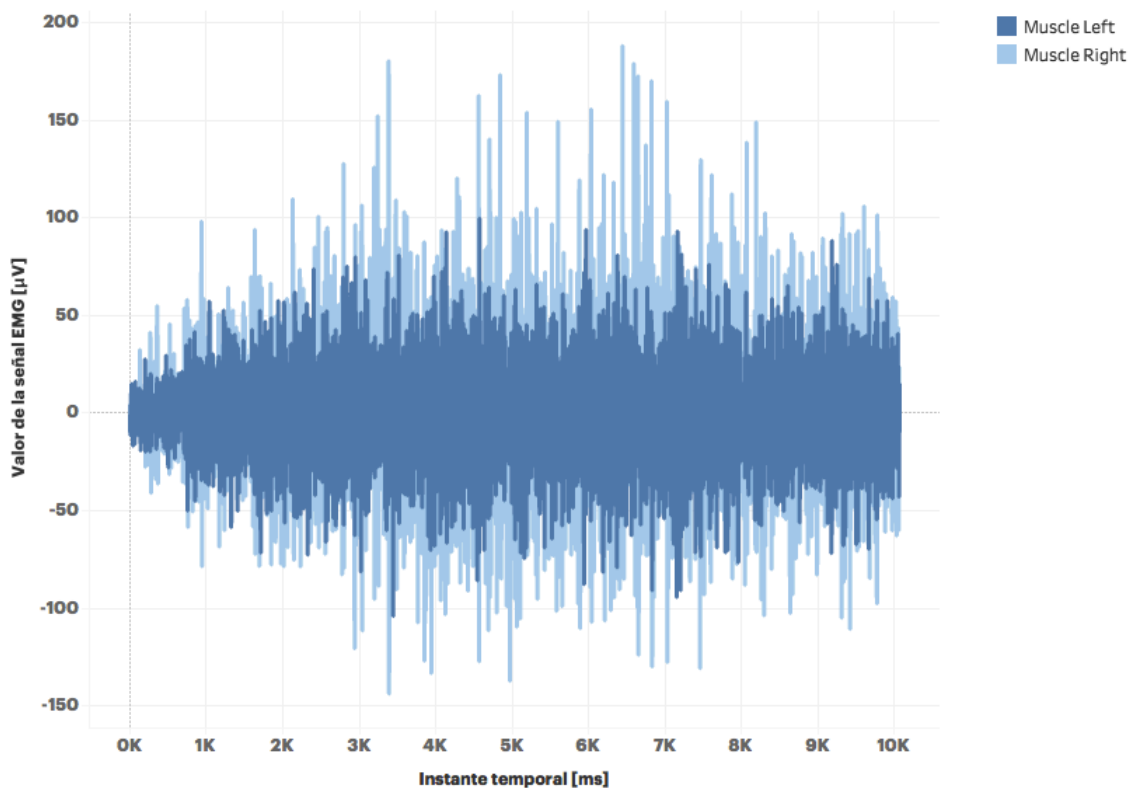


Figura 47. Señales correspondientes a los dos músculos cervicales una vez procesadas.

#### 4.5 Análisis de datos

Teniendo en cuenta lo anterior, si una señal EMG correctamente filtrada y limpia de ruido se asocia con la secuencia de movimientos FRP descritos anteriormente, se tendría un esquema muy parecido al representado en la figura 48 [50].

Cabe decir que aquí el nivel de señal EMG se encuentra normalizado y el FRP correspondería a un movimiento isotónico.

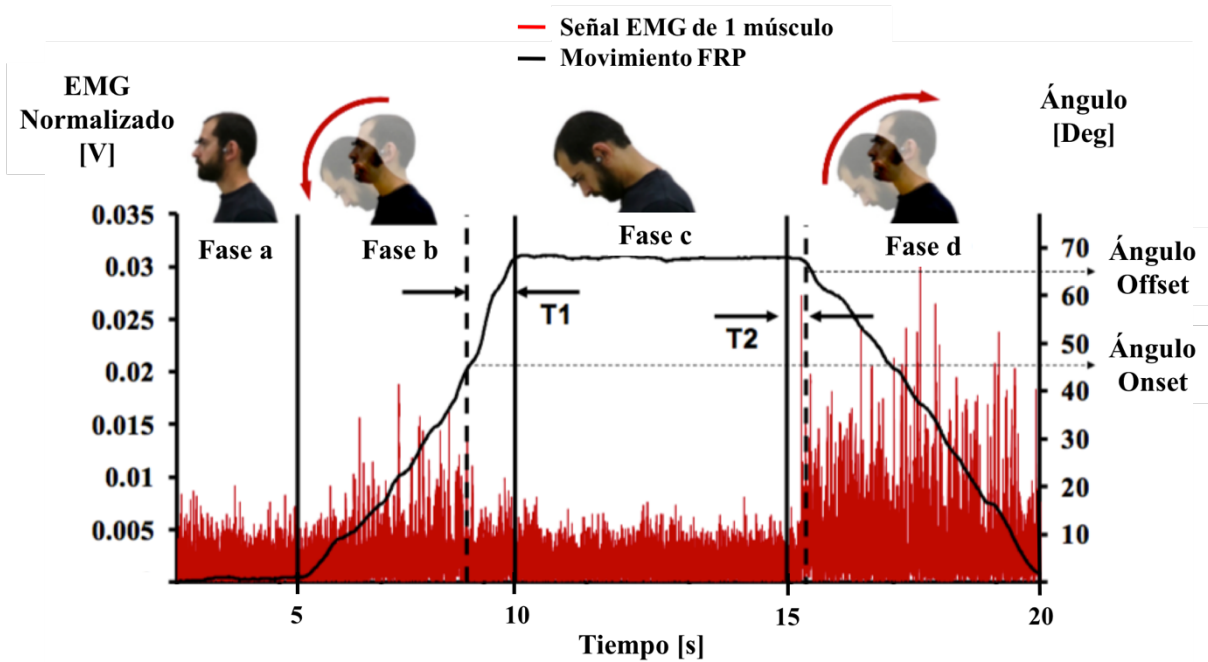


Figura 48. Esquema de equivalencia entre señal EMG y movimiento FRP de un test isotónico.

Así pues, la señal EMG posee algunas características muy interesantes y aprovechables desde el punto de vista sintáctico. Entre ellas:

- Un determinado **ratio de flexión-relajación (FRR)**, que mide la máxima actividad durante 1s de la completa extensión (fase d) y la compara con la actividad media en el proceso de flexión total (fase c). Este ratio cuantifica la capacidad de extensión de los músculos cervicales

$$FRR = \frac{EMG_{\text{máximo en Fase d}}}{EMG_{\text{promedio en Fase c}}} \quad (4.2)$$

- Dos ángulos distintos en función del nivel de fatiga de los músculos durante el cambio de fases en el movimiento FRP: **ángulos Onset y Offset**.
  - Por un lado, el ángulo Onset corresponde con el ángulo de flexión cuando el mayor nivel de carga justamente comienza a transferirse desde los músculos más activos (Fase b) hacia los menos fatigados (Fase c).
  - Por otro lado, el ángulo Offset corresponde con el ángulo de extensión mediante el cual el nivel de carga justamente termina de transferirse desde los músculos menos activos (Fase c) hasta los más activos (Fase d).

Estos ángulos en realidad se interpretan mediante el nivel instantáneo de flexión-relajación ( $IR_i$ ), donde:

$$IR_i = \frac{EMG_i}{EMG_{\text{máximo en Fase d}}} \quad (4.3)$$

Así pues, este nivel IRi define la proporción entre el nivel EMG instantáneo y el nivel EMG máximo en la Fase d. El instante en el que la flexión se encuentre por debajo del 40% del máximo se asociará con el nivel de tensión que en ese mismo instante proporciona el ángulo Onset; mientras que una proporción por encima del 40% del máximo se asociará con el nivel de tensión que en ese instante proporciona el ángulo Offset.

- Un **periodo de silencio** en la fase de extensión, que en realidad se define para cuantificar los cambios temporales producidos desde la Fase b hasta la c; y desde la Fase c hasta la d. O lo que es lo mismo, los instantes de tiempo que definen los ángulos Onset y Offset:

$$\text{Periodo de Silencio} = \text{Cambio Onset} + \text{Cambio Offset} \quad (4.4)$$

Donde, observando la figura 48, se definen:

$$\text{Cambio Onset} = T1_{\text{Fase c}} - T1_{\text{Fase b}} \quad (4.5)$$

$$\text{Cambio Offset} = T2_{\text{Fase c}} - T2_{\text{Fase d}} \quad (4.6)$$

Por tanto, atendiendo a estas últimas definiciones y a la de la ecuación 4.3, se puede decir que:

- El nivel EMG en el *instante 1 de la Fase b* es inferior al 40% del máximo en la Fase d, correspondiéndose este instante con el ángulo Onset.
- El nivel EMG en el *instante 2 de la Fase d* es superior al 40% del máximo en dicha fase, correspondiéndose este instante con el ángulo Offset.
- Un cálculo de suavizado, **Root Mean Square (RMS)**, que permite caracterizar datos procedentes de una señal de acuerdo a su contenido de energía.

$$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=0}^N x_i^2} \quad (4.7)$$

Este valor es tremendamente útil normalizado ya que permite comparar entre muestras de todo tipo, por lo que es interesante tener el nivel de señal en formato RMS y enventanado con 1 segundo de duración para poder hacer comparaciones visualmente más efectivas. Así pues, es común disponer del dato de RMS dividido por el total de segundos de la prueba (RMS/s) y también realizar una visualización porcentuada, ya que, de forma general, se podría decir que una persona muscularmente sana está por debajo del 10% en base a esa actividad.

- El **máximo valor de contracción muscular (MVC)** tanto de forma local como global, es decir, individualmente para cada prueba y también de forma global en todas las pruebas realizadas sobre un mismo músculo. Un factor interesante en este contexto procede de dividir un nivel de tensión por la máxima contracción

histórica, que coherentemente viene determinado en una prueba de movimiento isométrico.

#### 4.6 Etiquetado de datos

Ahora que ya se conocen las propiedades más importantes de las señales EMG y la forma en la que se ponen de manifiesto en la ejecución de movimientos FRP, se puede dar paso al análisis y justificación de la parte correspondiente al etiquetado de los conjuntos de datos que son candidatos a ser representados de forma gráfica.

Ya se ha comentado en la metodología propuesta sobre el capítulo anterior que el etiquetado de datos en nuestro experimento persigue una división entre características sintácticas y semánticas:

- Las *etiquetas sintácticas* proceden de los datos correspondientes a las señales obtenidas en las pruebas.
- Las *etiquetas semánticas* son las que se recogen en el experimento antes de comenzar la prueba de cada paciente mediante un breve cuestionario.

Así pues, como al fin y al cabo se busca una relación entre ambas, dichas etiquetas han de tener un nexo común que permita establecer relaciones y patrones visuales claros. Es por ello que el punto de unión entre etiquetas se lleva a cabo de la siguiente forma:

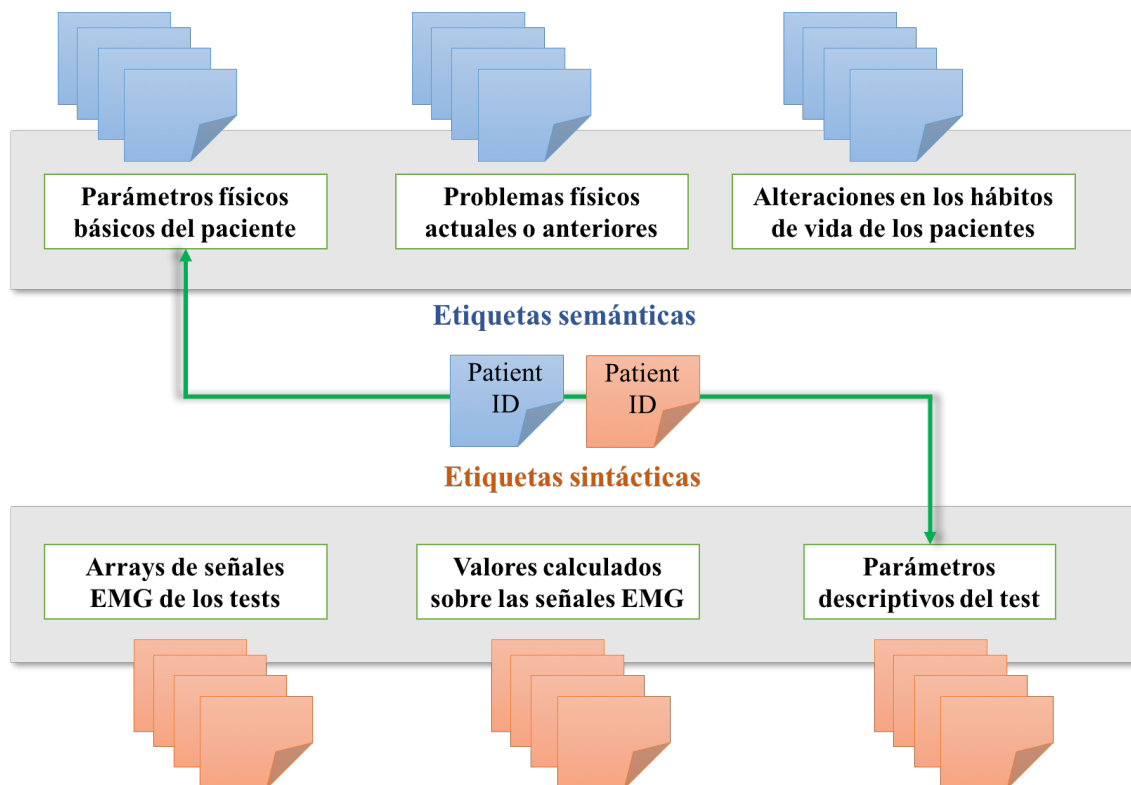


Figura 49. Esquema de mapeado entre etiquetas sintácticas y semánticas ejercido en el manejo de los datos.

Una vez tenido esto en cuenta, a continuación se procede a analizar las características semánticas y propiedades estructurales de los datos y gráficos que se pretenden manejar,

de modo que se apliquen posteriormente de forma relevante en el diseño del conjunto de herramientas de visualización personalizadas a estos datos.

Para facilitar la comprensión de todos los datos recogidos en el experimento, éstos se distribuyen a través de las tablas que se muestran a continuación.

#### 4.6.1 Etiquetas sintácticas

Poniendo atención en la figura 49 previamente mostrada, se aprecia que el campo de etiquetas sintácticas propuestas en este proyecto se divide en tres subcategorías claramente distinguibles, las cuales son objeto de análisis a continuación.

Por un lado, los **datos descriptivos del test** recogen información acerca de la prueba desde el punto de vista informativo. Realmente estos datos no suponen una amplia fuente de información, pero sí un punto de unión con las características semánticas.

Se muestra en la siguiente tabla el porqué:

Nombre etiqueta	Descripción	Ejemplo
id_patient	Identificador de paciente	292
date	Fecha en la que se realiza la prueba	Jaime García Giménez

Tabla 5. Modelo de datos descriptivo de las pruebas asociadas a un paciente.

Por otro lado, se recopilan las etiquetas consideradas de excelencia en nuestro experimento: los **arrays de datos con las propias señales EMG**.

Nombre etiqueta	Descripción	Unidades
time	Instante de tiempo de la prueba, a la cual se le establece un inventariado de 0,05 segundos.	[s]
ROM	Rango de movimiento, en grados, correspondientes a cada instante de tiempo	[°]
EMG_left	Nivel de señal EMG del músculo izquierdo correspondiente a cada instante de tiempo	[ $\mu$ V]
EMG_right	Nivel de señal EMG del músculo derecho correspondiente a cada instante de tiempo	[ $\mu$ V]
RMS_left	Nivel de señal suavizada del músculo izquierdo correspondiente a cada instante de tiempo	[ $\mu$ V]
RMS_right	Nivel de señal suavizada del músculo derecho correspondiente a cada instante de tiempo	[ $\mu$ V]
IR_left	Nivel de flexión-relajación para el músculo izquierdo	%
IR_right	Nivel de flexión-relajación para el músculo derecho	%

Tabla 6. Modelo de arrays de las pruebas asociadas a un paciente.

Por último, se etiquetan los **valores clave calculados** a raíz de los anteriores arrays, y que se utilizan para analizar estadísticamente las características sintácticas de cada una de las pruebas llevadas a cabo por los pacientes.

Nombre etiqueta	Descripción	Unidades
MVC_prueba_left	Máxima contracción muscular izquierda en la prueba	[ $\mu$ V]
MVC_prueba_right	Máxima contracción muscular derecha en la prueba	[ $\mu$ V]
MVC_historico_left	Máxima contracción muscular izquierda en todas las pruebas del paciente	[ $\mu$ V]
MVC_historico_rigth	Máxima contracción muscular derecha en todas las pruebas del paciente	[ $\mu$ V]
RMS_prom_left	Nivel de señal RMS promedio de la prueba para el músculo izquierdo	[ $\mu$ V]
RMS_prom_right	Nivel de señal RMS promedio de la prueba para el músculo derecho	[ $\mu$ V]
RMSsec_prom_left	Nivel de señal RMS promedio por segundo de la prueba para el músculo izquierdo	[ $\mu$ V]
RMSsec_prom_right	Nivel de señal RMS promedio por segundo de la prueba para el músculo derecho	[ $\mu$ V]
FRR_left	Capacidad de extensión del músculo izquierdo	[ $\mu$ V]
FRR_right	Capacidad de extensión del músculo derecho	[ $\mu$ V]
onset_angle_left	Ángulo onset del músculo izquierdo	[ $^{\circ}$ ]
onset_angle_right	Ángulo onset del músculo derecho	[ $^{\circ}$ ]
offset_angle_left	Ángulo offset del músculo izquierdo	[ $^{\circ}$ ]
offset_angle_right	Ángulo offset del músculo derecho	[ $^{\circ}$ ]
cambio_onset_left	Periodo de tiempo en el que se produce el ángulo onset en el músculo izquierdo	[s]
cambio_onset_right	Periodo de tiempo en el que se produce el ángulo onset en el músculo derecho	[s]
cambio_offset_left	Periodo de tiempo en el que se produce el ángulo offset en el músculo izquierdo	[s]
cambio_offset_right	Periodo de tiempo en el que se produce el ángulo offset en el músculo derecho	[s]
PS_left	Periodo de silencio del músculo izquierdo	[s]
PS_right	Periodo de silencio del músculo derecho	[s]

Tabla 7. Modelo de datos de valores calculados en las pruebas asociadas a un paciente.

#### 4.6.2 Etiquetas semánticas

Volviendo a fijarnos en la figura 49 podemos observar que el campo de etiquetas semánticas propuestas en este proyecto se divide en tres subcategorías bien diferenciadas, las cuales se definen a continuación.

En la base de datos de pacientes del experimento, se encuentran en primer lugar los **datos básicos de cada uno** de ellos, donde se recoge información de contacto del usuario; es decir, datos meramente organizativos:

Nombre etiqueta	Descripción	Ejemplo
id	Identificador de paciente	292
name	Nombre y apellidos de paciente	Jaime García Giménez
age	Edad de paciente	23
gender_display	Sexo de paciente	Hombre
email	Correo electrónico de paciente	<a href="mailto:jaimenig@ucm.es">jaimenig@ucm.es</a>
city	Ciudad de residencia del paciente	Madrid
occupation1	Profesión principal de paciente	Informático
occupation2	Segunda ocupación de paciente	Camarero
tests_count	Número de pruebas que el paciente ha realizado	3
height	Altura de paciente, en m	1,77
weight	Peso de paciente, en kg	72
IMC	Índice de masa muscular de paciente	22,98
physical_condition	Condición física promedio de paciente, en función del ritmo de vida con el que se autodefine (atlético, promedio o pobre)	Promedio

Tabla 8. Modelo de datos básicos de los pacientes del experimento.

En segundo lugar, se recogen los **problemas físicos de los pacientes**. Con esto se pretende determinar si el paciente posee algún tipo de dolencia física o está completamente sano. Para ello, se crea una etiqueta específica (*physical\_problems*) que es la que define automáticamente si un paciente se encuentra (o no) totalmente sano en función del resto de variables.

Esta etiqueta altamente restrictiva ya que, en caso de tener un solo valor no positivo en alguna de las variables que le acompaña, ésta automáticamente también pasa a tener un valor negativo para ese paciente. Este tipo de exclusión es común para estimar los casos peores entre pacientes y poder investigar causas.

En la tabla 9 se recogen estas etiquetas:

Nombre etiqueta	Descripción	Rango	Ejemplo
physical_problems	Problemas físicos de los pacientes. Etiqueta de	<ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>	No ( <i>Nótese que este caso únicamente se</i>

	exclusión en función de las siguientes		<i>cumpliría bajo los valores indicados por las siguientes etiquetas)</i>
muscular_fatigue	Nivel de fatiga muscular de paciente	<ul style="list-style-type: none"> <li>• 0 (Sin fatiga muscular)</li> <li>• 1 (Fatiga muscular leve)</li> <li>• 2 (Fatiga muscular intermedia)</li> <li>• 3 (Fatiga muscular intensa)</li> </ul>	0
pain_level	Nivel de dolor en los músculos cervicales de paciente	<ul style="list-style-type: none"> <li>• 0 (Sin dolor)</li> <li>• 1 (Dolor cervical leve)</li> <li>• 2 (Dolor cervical intermedio)</li> <li>• 3 (Dolor cervical intenso)</li> </ul>	0
injured	Determina alguna lesión en el paciente	<ul style="list-style-type: none"> <li>• Escoliosis</li> <li>• Hernia</li> <li>• (Otros)</li> <li>• Posible (En caso de encontrarse bajo pruebas médicas)</li> <li>• No</li> </ul>	No
observations	Observaciones que supongan algún tipo de impedimento físico	<ul style="list-style-type: none"> <li>• Artrosis</li> <li>• Bajo pruebas médicas</li> <li>• Esguince cervical</li> <li>• (Otros)</li> <li>• No</li> </ul>	No
sleep_fatigue	Cansancio tras dormir	<ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>	No
special_diet	Condiciones alimentarias especiales	<ul style="list-style-type: none"> <li>• Diabético</li> <li>• Celiaco</li> <li>• Vegetariano</li> <li>• (Otros)</li> <li>• No</li> </ul>	Celiaco
others	Otras dificultades o peculiaridades en los hábitos de vida del paciente	<ul style="list-style-type: none"> <li>• Náuseas</li> <li>• Mareos</li> <li>• Sensibilidad a la luz</li> <li>• Migrañas</li> <li>• (Otros)</li> <li>• No</li> </ul>	Mareos

Tabla 9. Modelo de problemas físicos de los pacientes del experimento.

Por último, se definen las etiquetas encargadas de definir **cualquier tipo de alteración** que pueda influir en el día a día de un paciente. Para ello, se sigue el mismo proceso que en el anterior modelo de datos. Se establecen una etiqueta específica (*alterations*) que es la que define automáticamente si un paciente posee (o no) cualquier tipo de dificultad entre sus hábitos de vida cotidianos.

Esta etiqueta también es altamente restrictiva ya que se negativiza con solamente una de las que le definen, por la misma razón comentada anteriormente.

El conjunto de etiquetas que definen estas características de población se definen comentan sobre la tabla número 10:

Nombre etiqueta	Descripción	Rango	Ejemplo
alterations	Alteraciones en los hábitos de vida de los pacientes. Etiqueta de exclusión en función de las siguientes	<ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>	No ( <i>Nótese que este caso únicamente se cumpliría bajo los valores indicados por las siguientes etiquetas</i> )
ergonomic_problems	Problemas ergonómicos sufridos por el paciente	<ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>	No
stress_level	Situación de estrés general	<ul style="list-style-type: none"> <li>• Agudo puntual</li> <li>• Agudo episódico</li> <li>• Estrés crónico</li> <li>• Ninguno</li> </ul>	Ninguno
sleep_fatigue	Cansancio tras dormir	<ul style="list-style-type: none"> <li>• Sí</li> <li>• No</li> </ul>	No
special_diet	Condiciones alimentarias especiales	<ul style="list-style-type: none"> <li>• Diabético</li> <li>• Celiaco</li> <li>• Vegetariano</li> <li>• (Otros)</li> <li>• No</li> </ul>	No
others	Otras dificultades o peculiaridades en los hábitos de vida del paciente	<ul style="list-style-type: none"> <li>• Náuseas</li> <li>• Mareos</li> <li>• Sensibilidad a la luz</li> <li>• Migrañas</li> <li>• (Otros)</li> <li>• No</li> </ul>	No

Tabla 10. Modelo de datos de alteraciones en los hábitos de vida de los pacientes del experimento.

#### 4.7 Optimización de la plataforma, estado to-be

Como bien sabemos, la propuesta esencial del estudio práctico es potenciar los modelos de datos y gráficos contemplados en la herramienta as-is de mDurance en base a los modelos de informes susceptibles a adaptar y recuperar en la metodología CBR general de recuperación de plantillas de informes explicada sobre el capítulo anterior.

Aquí, el sub-sistema CBR es el encargado de elegir las técnicas visuales que mejor se adapten mejor al hueco del informe, de modo que se puedan ligar las etiquetas sintácticas y semánticas del experimento en base a las entradas de datos proporcionadas y crear gráficos que proporcionen valor sobre estos conjuntos.

Por un lado, la *caracterización de datos* comentada en el capítulo 4.6, la cual sostiene un lenguaje común en la descripción de los distintos huecos integrados en la plantilla de

informes. Son precisamente estos huecos los que se pretenden rellenar y los que dependen de forma intrínseca del usuario encargado de proporcionar las entradas de datos, que es el mismo al que se dirige el informe y que será el espectador objetivo de la utilización de nuestros elementos visuales.

Por otro lado, la *caracterización de los gráficos* candidatos a ser aplicados en el hueco del informe en función de las reglas genéricas comentadas en el apartado 3.4 de la metodología, y con la debida particularización que se comenta sobre el siguiente apartado 4.9. Estos gráficos han de poseer la heterogeneidad suficiente como para establecer un repertorio que suponga una notable mejora en la recuperación y métodos de reutilización de un ciclo CBR para así permitir la combinación de distintos elementos visuales en un mismo caso base. Para esta combinación se ha de tener muy en cuenta la implementación de un formato único en ciertas partes de los elementos visuales independientemente del hueco de informe que se desee completar, con tal de no crear confusión en el usuario que acostumbra a visualizar gráficos en este tipo de informes.

#### 4.7.1 Modelos de informes propuestos

El enfoque capaz de potenciar la herramienta de mDurance en una integración CBR se podría adecuar perfectamente a los prototipos de informes que cubren las necesidades de los siguientes ficheros de entrada:

1. **Informe para el estudio de la población de pacientes:** Nuestra fuente de datos es una base de datos con las características de dicha población y unos atributos determinados.
2. **Informe para el estudio del estado físico de un paciente:** Nuestra fuente de datos es un fichero donde se recoge la actividad muscular de un solo paciente.
3. **Informe para la comparación de dos pacientes:** Nuestra fuente de datos son dos ficheros donde se recoge la actividad muscular de cada uno de los pacientes.
4. **Informe para la comparación de un paciente con respecto a la población:** Nuestra fuente de datos es un fichero de la actividad muscular de un solo paciente y otro con los datos promedios de la población.

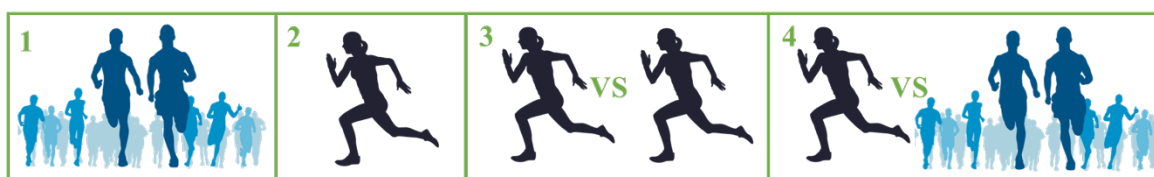


Figura 50. Propuesta de modelos de informes para la reutilización y adaptación de gráficos.

#### 4.7.2 Ejemplo general de utilización de los informes

De este modo, a continuación vamos a suponer un ejemplo en el que se tenga como base el ciclo CBR general que tenga como objeto la recuperación de los modelos de informes previamente descritos, para así utilizar nuestro sub-sistema CBR en la reutilización y adaptación de gráficos en base a experiencias previas.

Una clínica de fisioterapia cuenta con la herramienta de salud digital de mDurance, donde se almacena una base de datos de pacientes y otra de pruebas, las cuales se van

actualizando de forma constante. Además, también se tiene un histórico de las fuentes de datos que proporcionan los sistemas Shimmer3 que se tienen en la clínica.

En caso de requerir un estudio de la población de pacientes que se tiene en la clínica, mediante la opción de solicitud explícita en el sistema se podría conseguir este tipo de informe, que recoge los principales datos de la población en función de las necesidades que indique el experto.

Supongamos que un nuevo paciente, miembro de un equipo de natación sincronizada, por ejemplo, ingresa en la clínica y realiza las pruebas EMG. Se puede dar el caso de que este perfil de deportista acabe de salir de una lesión o que simplemente quiera revisar el estado de sus músculos cervicales. En este caso, lo más común es que el experto médico requiera al sistema de un informe de dicho paciente donde aparezcan sus características físicas y gráficos que representen los valores descriptivos de las pruebas EMG que ha realizado, para ver los niveles neuromusculares que implicados y la evolución del rango de movimiento. Esto permite, por un lado, saber si se ha realizado correctamente la prueba debido al previo establecimiento de valores críticos en los márgenes; y por otro, también permite detectar anomalías neuromusculares y comprobar la distribución de carga de los músculos del paciente.

No obstante, el papel del experto médico no termina aquí. Supongamos que además también pretende comparar los niveles de este paciente con los de otro componente de su mismo equipo de natación sincronizada. Pues bien, el experto médico solicita la información comparativa de los dos pacientes al sistema para de este modo reutilizar gráficos por parte de los dos pacientes comparándolos entre sí y destacando las diferencias notables en sus niveles musculares, pudiendo también reconocer patrones entre este mismo perfil de usuarios y categorizarlos en base a una serie de anomalías concretas.

Por último, para constatar los datos tomados del deportista de natación sincronizada, el experto médico pretende comparar este perfil de paciente con otros usuarios, concretamente con la población en su totalidad. Pues bien, el ejercicio que realiza el experto médico en este aspecto es una solicitud al sistema para comparar este paciente en particular con el resto de la población, y así contemplar patrones que relacionen (o no) este paciente con otros perfiles diversos. Estos patrones no tienen el mismo nivel de abstracción que la anterior comparación paciente a paciente, sino que son de carácter poblacional y entra en juego un abanico de datos mucho más contextual.

Todo lo anterior forma parte del ciclo CBR general en el que se recuperan los informes y se generan los conjuntos de datos con la información necesaria para ello. Ya se ha dicho que este Trabajo Final de Máster no se centra en ese ciclo general pero, no obstante, para la reutilización y adaptación de gráficos sí hace falta al menos partir de un supuesto caso en el que se tengan ciertos huecos para completar con los elementos visuales reutilizados. Es por ello que en la siguiente figura se propone una esquematización básica de las formas en las que el experto médico puede hacer una petición para así generar los datos correspondientes, y de paso el sub-ciclo CBR comenzar a tantear los gráficos candidatos a ser reutilizados.

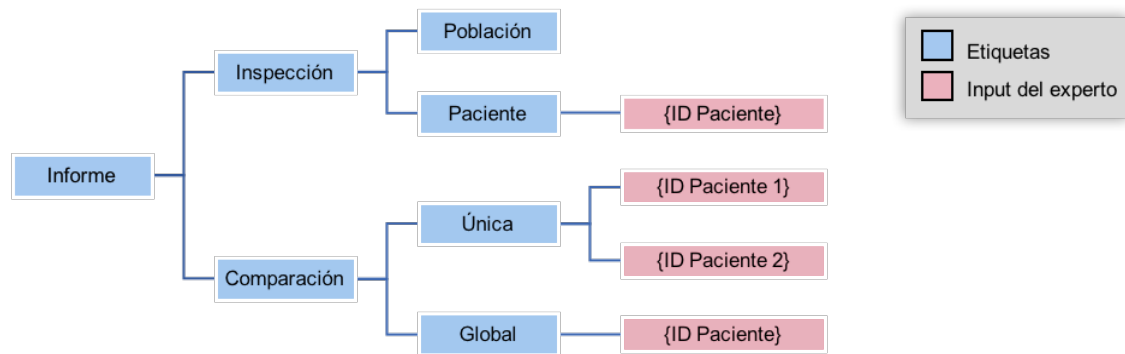


Figura 51. Etiquetas de lenguaje fisioterapeuta-sistema para la suplantación del ciclo CBR general

De este modo, si por ejemplo nos centramos en el caso en el que el fisioterapeuta quiere saber el estado físico de un paciente tras la prueba EMG, tendría que recurrir a una inspección de un paciente en concreto, introduciendo el identificador de dicho deportista. Así pues, se tendrían las etiquetas: *Informe*, *Inspección*, *Paciente*, *{ID Paciente}*.

#### 4.8 Diseño de elementos visuales objeto de reutilización y adaptación

Para finalizar este capítulo se procede a describir el diseño e implementación de los elementos DataViz fruto del experimento y, por consiguiente, del presente Trabajo Final de Máster. Éstos componen la base de casos dentro del sub-ciclo CBR para la reutilización y adaptación de los gráficos según la entrada de datos determinada por la petición del experto médico.

Para ello, se analiza y justifica el diseño e implementación de los elementos visuales propuestos para cada tipo de informes. Así pues, se tiene en cuenta todo lo comentado en el apartado 3.4 de *caracterización de gráficos* además de una serie de reglas orientadas al objetivo de este caso de estudio, con el principal propósito de convertir uno o varios gráficos en una herramienta DataViz capaz de proporcionar valor en nuestro contexto de pruebas EMG en la Medicina del Deporte.

Algunas de estas reglas se basan en:

- Acompañar cada una de las herramientas visuales con una breve y clara descripción, por lo que se hace uso de tablas que muestren el valor de una variable acompañado del nombre que la define.
- Indicar en todos los ejes las dimensiones de las etiquetas sintácticas que se representan.
- Utilizar ciertas dimensiones propias de la población con una única paleta de colores bien distinta a la que caracteriza las comparaciones entre otros factores. Las paletas de colores empleadas se recogen en el anexo 2 de este documento.
- Dotar a todos los gráficos representados a de carácter interactivo. Esto es, permitir al experto médico que consiga un mayor nivel de detalle acerca de determinadas medidas al posicionar el cursor sobre los gráficos, y permitir la opción de filtrado para concretar en las dimensiones de interés.

#### 4.8.1 Estudio de la población de pacientes

Para el primer tipo de informes se proponen distintas herramientas DataViz encargadas de cubrir y relacionar el mayor número de etiquetas entre las recogidas en el experimento. Ya se ha visto que, además de características físicas, muchas de estas etiquetas contextuales son cadenas de texto que definen a los pacientes, por lo que para este tipo de informes las reglas de similitud son las siguientes:

1. Se destaca el uso de etiquetas nominales. Sin ir más lejos, la etiqueta que se corresponde con el *sexo* de paciente supone de gran importancia ya que en este contexto existen distintas formas de comportamiento en las dolencias cervicales entre hombres y mujeres. Una buena forma de caracterizar gráficamente estas etiquetas nominales es mediante el color o simplemente mediante texto.
2. En caso de querer obtener porcentajes, se utilizan gráficos de pastel que logren relacionar la dimensión con las proporciones dentro de una misma variable.
3. Para distribuciones jerárquicas, como causas de alteraciones en los hábitos de vida de los pacientes o problemas físicos, se utilizan mapas de árbol o diagramas de dispersión/burbujas donde se observe la cantidad de pacientes que poseen las características propias de desorden o dolencia.
4. En caso de representar la ubicación del paciente necesariamente se recurre a un mapa geográfico.

Uno de los puntos más generales que destacan entre este tipo de informes es la descripción global de la propia población, que es fácilmente representable en forma de proporciones y con el apoyo de etiquetas de texto donde se expongan de forma clara y concisa los datos y porcentajes más destacables.

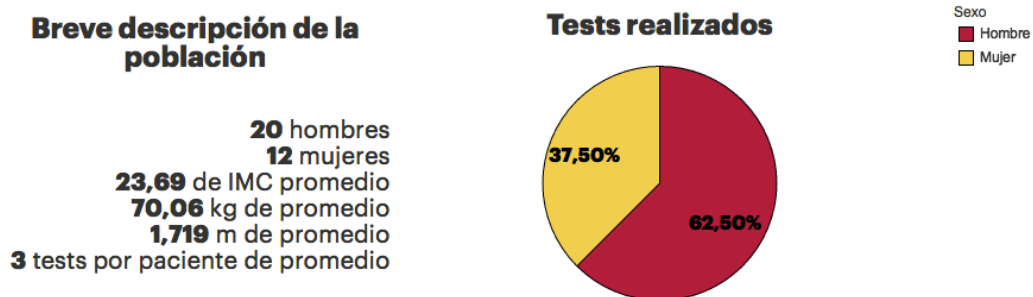


Figura 52. Breve descripción de la población y proporción de tests realizados.

Además, existen determinadas características básicas que determinan la condición física de un paciente dentro de la población, como bien son la estatura, el peso, y por consiguiente el IMC. Esto son datos cuantitativos, y el mensaje que se quiere transmitir es de comparación a modo de ranking, por lo que se representan mediante un diagrama de barras y se aprovecha la etiqueta de *condición física* como ejes de filas.

Así pues, al no poseer un elevado número de tipos de condición física, la visualización de las barras queda sujeta a ser clara y concisa.

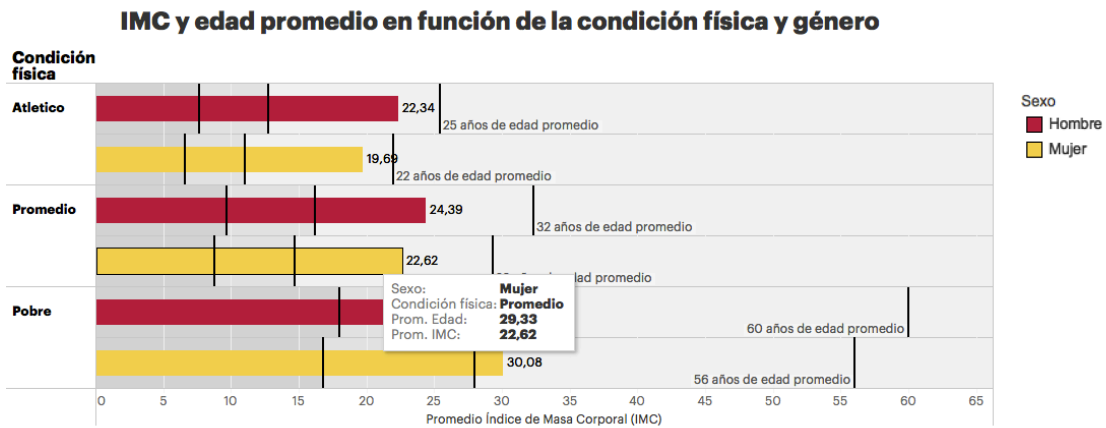


Figura 53. Representación del IMC y edad promedio en función de la condición física y género de los pacientes.

Por otro lado, mediante un mapa geográfico se puede mostrar la ciudad de residencia de los pacientes. Realmente esto depende mucho de la envergadura que se quiera imponer en el sistema, pero de cualquier forma resulta de interés una vez disponemos de la etiqueta de *ciudad*.



Figura 54. Representación de la geolocalización de los pacientes de la población.

También se puede saber más sobre la relación entre las etiquetas de *altura* y el *peso* de los pacientes, estableciéndolos en los ejes cartesianos de modo indiferente y asociándolos con la *edad* para saber la forma de distribución y regresión lineal que siguen estos atributos físicos, ahora independientes del nivel de IMC.

El objetivo de representar la etiqueta de *edad* en este gráfico es simplemente para tanteear la población de un vistazo. Por ejemplo, se podrían observar x casos de pacientes relativamente jóvenes que se encuentran totalmente alejados de la recta de regresión lineal: muy por debajo del nivel promedio de altura y por encima del nivel promedio de peso.

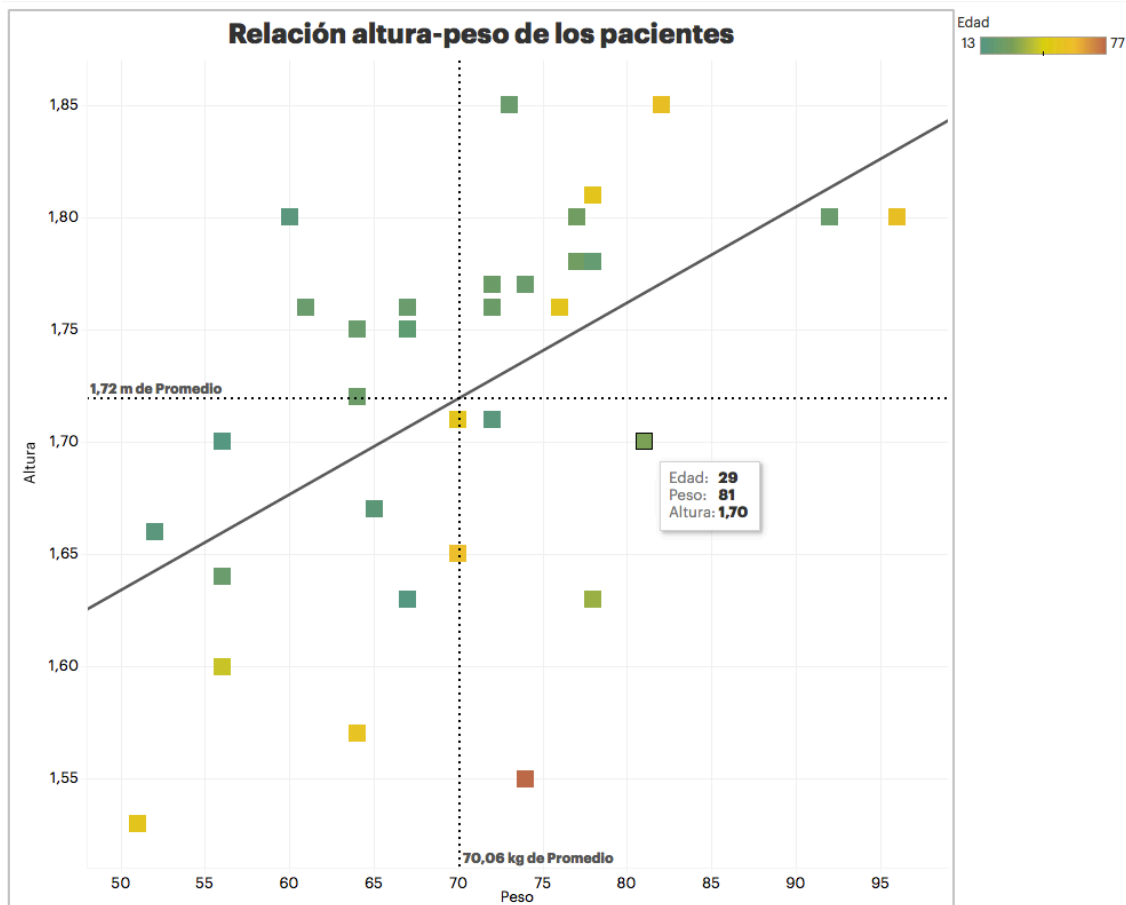


Figura 55. Distribuciones y recta de regresión de la relación altura-peso en función de la edad de los pacientes.

Una vez aquí, se considera de especial interés el nivel de interacción del usuario (o experto médico) con la herramienta, ya que se establecen filtros donde se muestra el interés del mismo con cada uno de las medidas que ofrece la visualización.

La siguiente figura muestra una de las formas de contemplar esto, y en este caso se hace relacionando la distribución de pacientes que muestra un determinado *nivel de estrés* en sus hábitos de vida y también si sufre de *dolores cervicales*. Contextualmente, estas dos etiquetas son tremendamente útiles para observar la cantidad y promedio de pacientes que se encuentran entre los máximos y mínimos poblacionales y la relación que poseen entre sí.

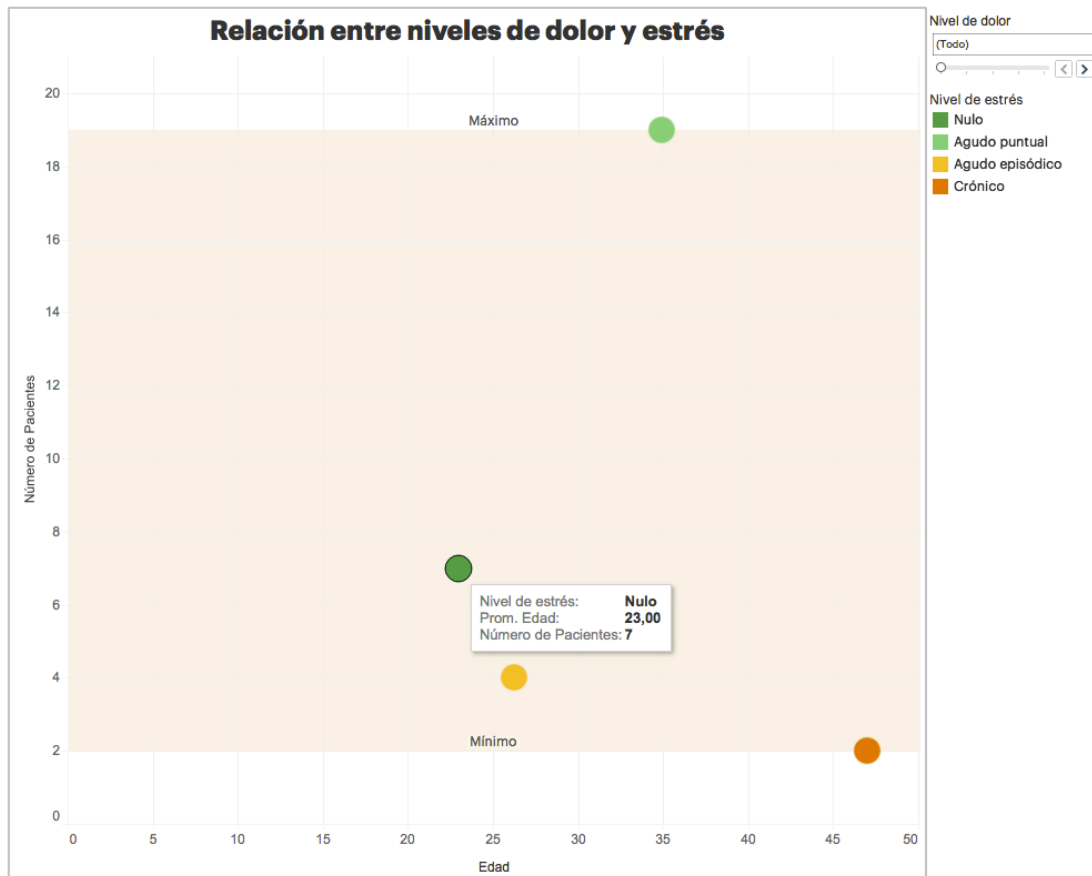


Figura 56. Representación de la relación entre nivel de dolor y estrés en la población de pacientes.

Se utiliza un gráfico de dispersión en el que los ejes indican el *número de pacientes* y la *edad* de los mismos. Si nos fijamos, una alternativa sería utilizar el atributo de tamaño para las circunferencias y prescindir de los ejes, lo cual convertiría este gráfico en un diagrama de burbujas. No obstante, en este caso perderíamos la etiqueta de la edad, que juega un papel importante precisamente en este contexto.

Se observa a continuación el reajuste mediante la opción de filtrado con dos opciones posibles:



Figura 57. Filtrado en la representación de la relación entre nivel de dolor y estrés.

Otro tipo de gráficos que permiten un alto nivel de interacción son los diagramas de árbol, donde ya no disponemos de ejes cartesianos como en el caso anterior, sino que se juega con el atributo del tamaño de los bloques. Esto es posible ya que el mapa de árbol y el conjunto de áreas que lo componen son suficientes como para albergar tres o cuatro etiquetas de texto por bloque, evitando así los ejes cartesianos para segregar la información. Se tiene una cantidad de etiquetas (nominales) mayor, en un espacio mucho más optimizado y compacto.

Estos gráficos permiten dos opciones más que interesantes en este contexto:

1. Contemplar las causas de algún tipo de alteración en los pacientes, mediante el análisis de los atributos que hacen que los pacientes puedan poseer algún desorden; esto es, el *nivel de estrés*, los *problemas ergonómicos*, algún tipo de *alimentación especial*, o *desorden del nivel de sueño*.

### Causas de las alteraciones entre los hábitos de vida de los pacientes



Figura 58. Representación de las causas de las alteraciones en los pacientes mediante mapa de árbol.

2. Contemplar las causas de algún tipo de problema físico en los pacientes, mediante las etiquetas que hacen que los pacientes puedan poseer alguna dolencia; esto es, el propio *nivel de dolor cervical*, la *fatiga muscular*, o algún tipo de *lesión* concreta.

## Causas de los problemas físicos de los pacientes

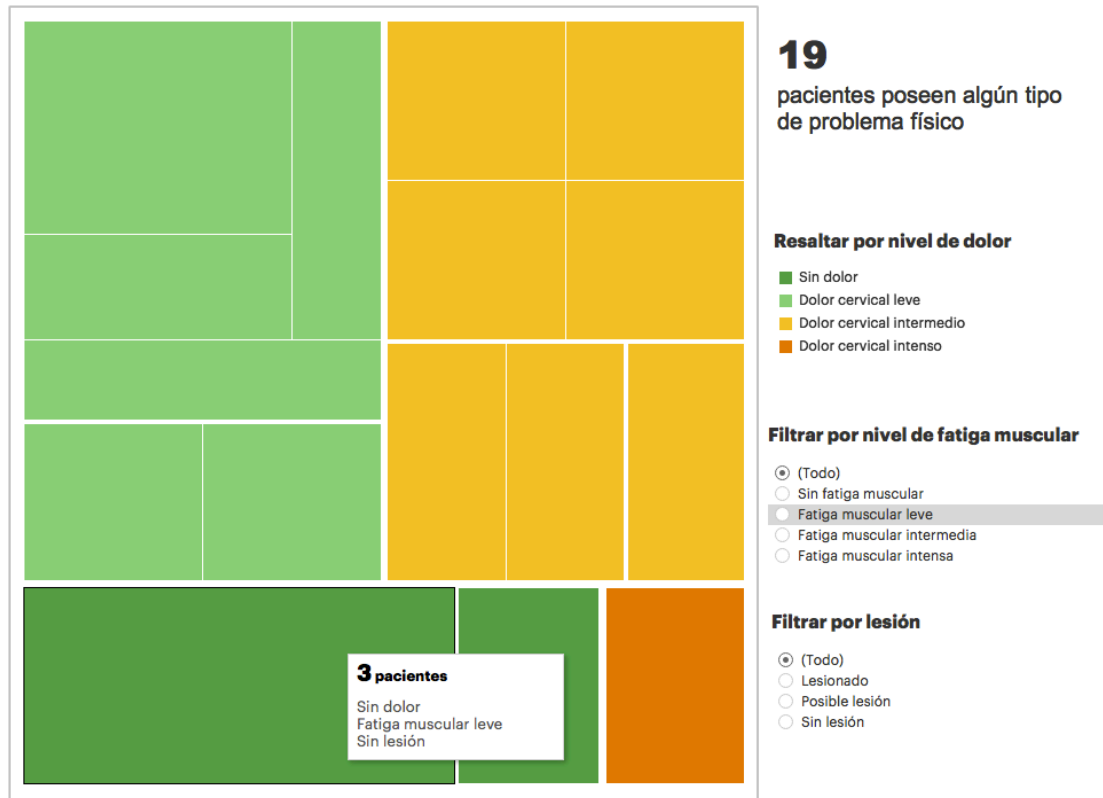


Figura 59. Representación de las causas de los problemas físicos en los pacientes mediante mapa de árbol.

El filtrado o exclusión de características que hacen que un paciente (o grupo de pacientes) posea alteraciones en sus hábitos de vida o algún tipo de dolor en las cervicales, convierte el mapa de árbol en un bloque explicativo con más características comunes.

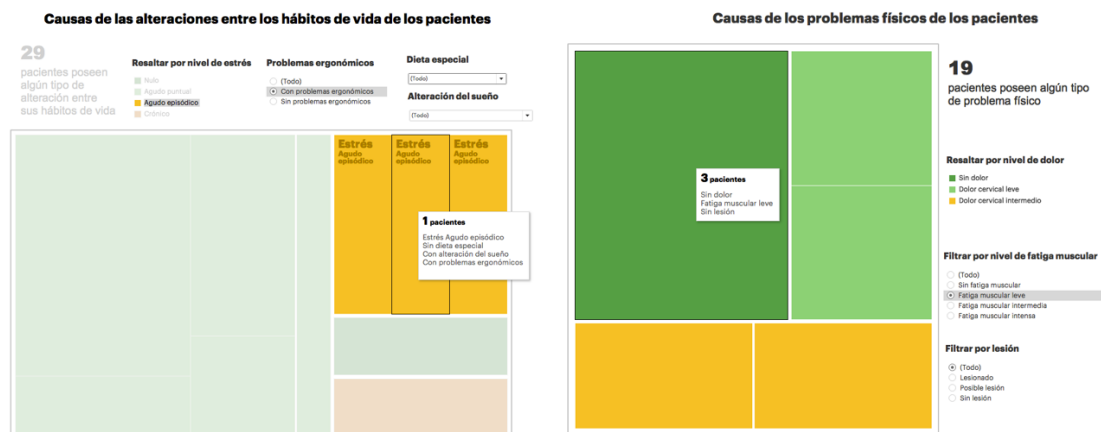


Figura 60. Filtrado en la representación de causas en alteraciones y problemas físicos de los pacientes.

Por último, también es importante para un experto médico contemplar la profesión de los pacientes en la tenencia (o no) de problemas ergonómicos. Así pues, una buena

forma de representar esto es a través de dos clústers de pacientes donde asociar estas características comunes a un campo donde influyen distintos niveles de dolor cervical y de estrés a los nombrados problemas ergonómicos entre pacientes con distintas ocupaciones. Este gráfico multietiqueta no es totalmente trivial y una de las pocas opciones que deja es el diagrama de burbujas:

### Influencia de la ocupación y los problemas ergonómicos de los pacientes con su nivel de dolor

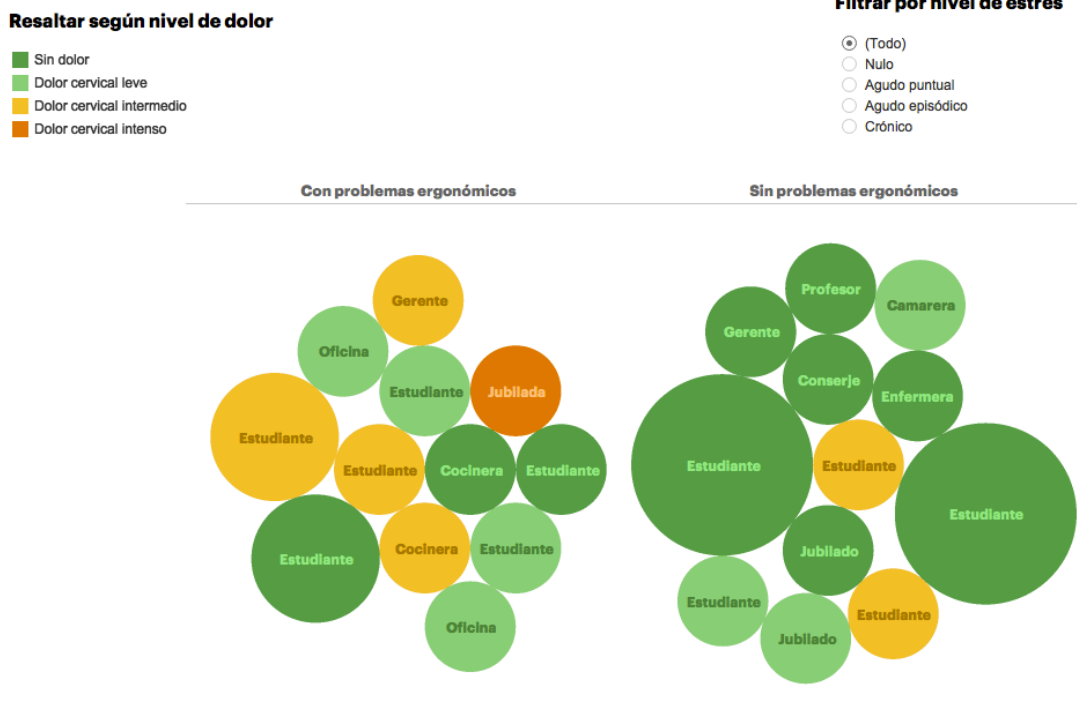


Figura 61. Representación de la ocupación en relación con los niveles de estrés y dolor cervical de los pacientes.

De nuevo, filtrando el gráfico, el usuario puede centrarse en la etiqueta de *profesión* que más interés cause en el contexto.

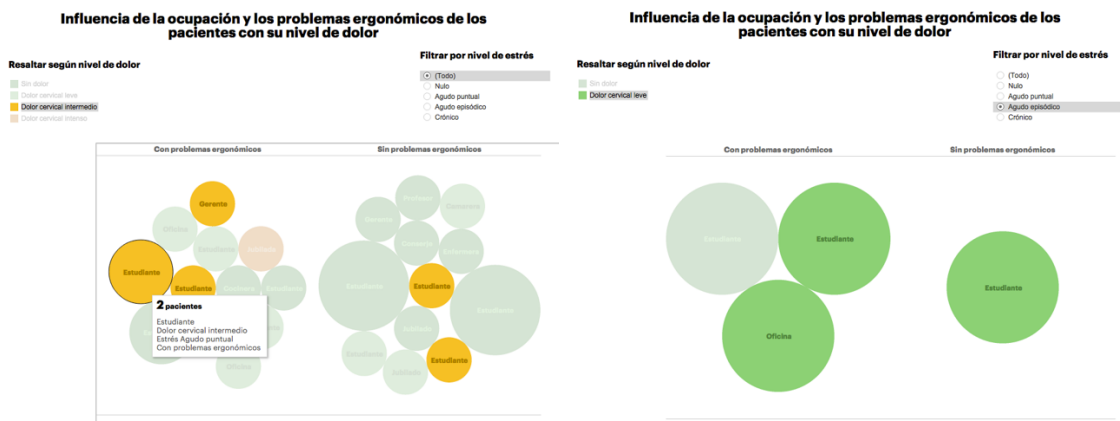


Figura 62. Filtrado en la representación que relaciona la ocupación de los pacientes con los niveles de estrés y dolor cervical.

#### 4.8.2 Estudio del estado físico de un paciente

El segundo tipo de informes, al contrario que los anteriores, se basa casi de forma completa en las etiquetas sintácticas en lugar de las semánticas, pues se corresponden con las pruebas que un único paciente ha ejercitado.

En este tipo de estudio priman las métricas en lugar de las dimensiones, y es por ello que se tienen de forma mucho más priorizada el uso de valores puntuales, evoluciones y distribuciones. De este modo, las reglas de similitud establecidas para este tipo de informes son las siguientes:

1. Para representar valores cuantitativos particulares o niveles atípicos, tanto los gráficos de barras como los de dispersión son buenos candidatos. Únicamente se ha de tener en cuenta la cantidad de elementos que componen el gráfico ya que, en caso de ser elevada, un diagrama de dispersión sería más interesante pues no satura tanto al visualizador, ya que con el diagrama de barras influyen los colores en mayor medida.
2. Las evoluciones sobre el tiempo de dos músculos se muestran mediante gráficos de líneas, de modo que se pueda observar y comparar la tendencia temporal entre ellos, además de generar valores estadísticos sobre la relación de forma y tamaño en el tiempo.
3. Para los casos en los que únicamente se tenga en cuenta el nivel de distribución independientemente del tiempo, se puede hacer provecho de los gráficos de caja y bigotes y su capacidad para mostrar de manera estadística la carga muscular del paciente en cuestión.

De forma general, se podría hacer visible un resumen de las características de dicho paciente de la forma en la que se expresa en la figura 63, donde se representa una *descripción breve* sobre sus parámetros básicos, el *nivel de ejercicio promedio* de cada músculo en las tres pruebas llevadas a cabo, y la *fecha* de realización de cada una de estos *tres tests*.

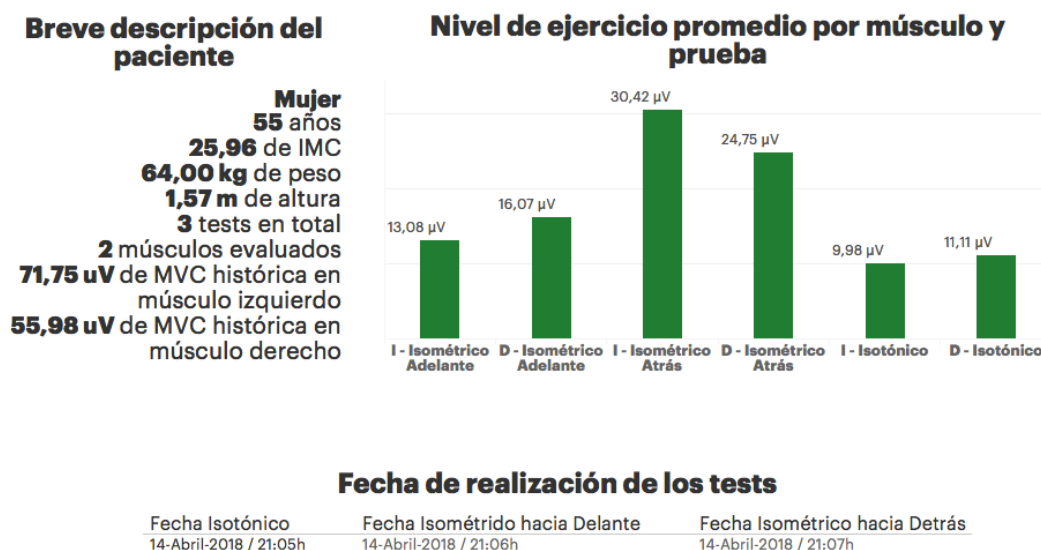


Figura 63. Descripción, ejercicio promedio y resultados destacables de las pruebas realizadas por el paciente.

Realmente las opciones para completar los tipos de huecos de este tipo de informes son múltiples, pues depende de tantas pruebas como el paciente haya realizado. En nuestro caso, procedemos a analizar los gráficos de una de ellas, la del movimiento isotónico, pues es la que mayor duración posee y cuyas señales mejor representan el movimiento FRP.

Esto se puede comprobar, sin ir más lejos, en el *rango de movimiento* de esta prueba, donde se pueden ajustar parámetros estadísticos básicos que ayuden a comprender el estado de la prueba y detectar anomalías en la ejecución de la misma. En este contexto, hay ocasiones en la que los pacientes exceden el movimiento FRP hasta un nivel determinado y por tanto la prueba no podría considerarse como válida.

Este diagrama del rango de movimiento se representa mediante un gráfico de líneas casi por defecto, ya que para los expertos resulta importante observar la forma de la secuencia llevada a cabo por el paciente en la prueba. Es por ello que el fondo que determina la línea y los ejes se le adjudica un tono de color vistoso que lo diferencie del fondo.

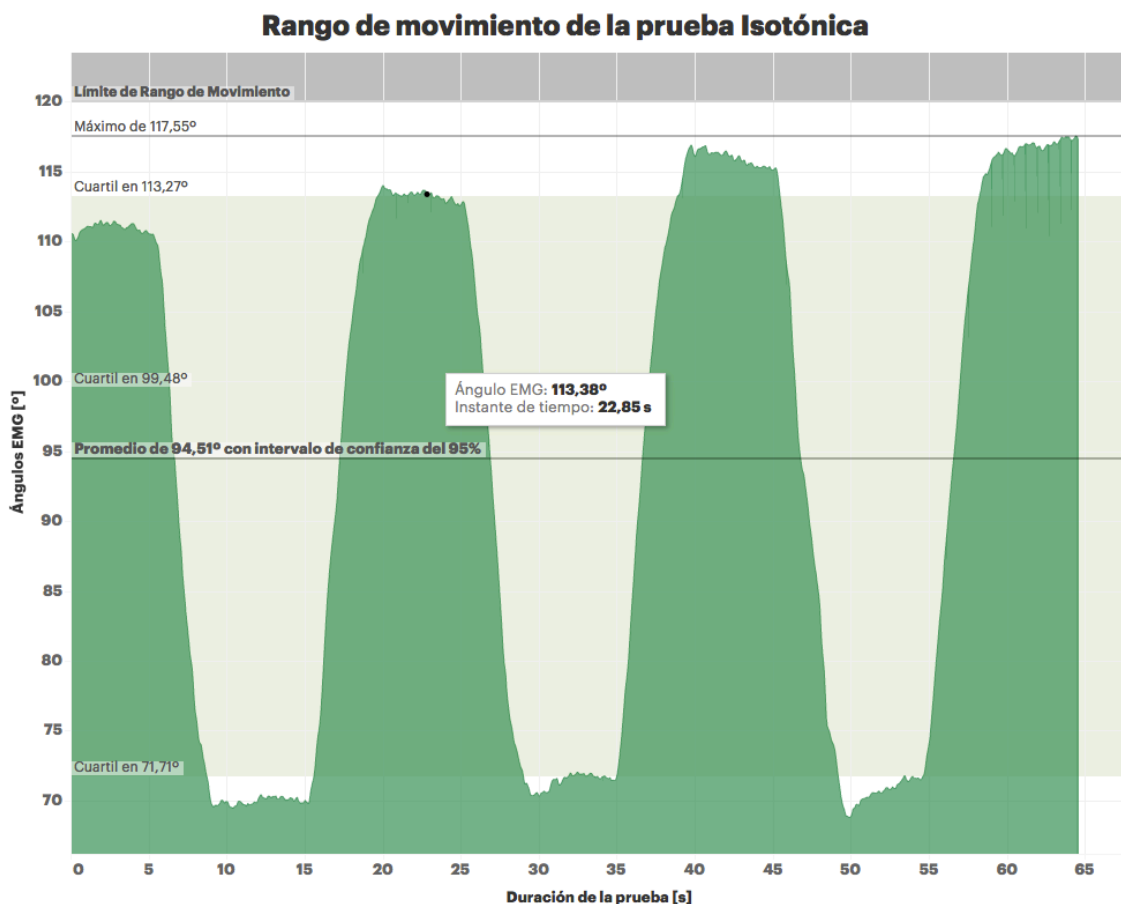


Figura 64. Representación del Rango de Movimiento del paciente con cálculos estadísticos de distribución.

Además, otro factor analítico de interés en este contexto es la distribución del *ratio de flexión-contracción instantáneo*, o lo que es lo mismo, la cantidad de ejercicio que realiza cada uno de los dos músculos de forma instantánea en comparación con el máximo realizado tras la fatiga muscular.

Este concepto puede servir de ayuda para comprobar anomalías como la que se observan en la figura 65, donde a pesar de tener un nivel promedio del 50,3%, el paciente posee una distribución de ratios totalmente desigualada, destacando el músculo izquierdo totalmente sobre el derecho.

Una vez más, se utiliza el gráfico de líneas, pero esta vez complementado perfectamente con descripciones como las mostradas abajo, de modo que se desglosen numéricamente los más importantes tomados en la prueba para así servir de ayuda para el usuario especialista.

### Rendimiento del músculo cervical izquierdo en la prueba Isotónica

**7,64 uV** de FRR  
**13,91 uV** de RMS promedio  
**0,22 uV/s** de RMS promedio  
**31,70 uV** de MVC máximo en la prueba  
**75,41°** de ángulo onset  
**104,61°** de ángulo offset  
**3,11 s** de cambio onset  
**0,56 s** de cambio offset  
**3,67 s** de periodo de silencio

### Rendimiento del músculo cervical derecho en la prueba Isotónica

**4,50 uV** de FRR  
**9,56 uV** de RMS promedio  
**0,15 uV/s** de RMS promedio  
**9,56 uV** de MVC máximo en la prueba  
**103,58°** de ángulo onset  
**105,05°** de ángulo offset  
**1,12 s** de cambio onset  
**0,51 s** de cambio offset  
**1,63 s** de periodo de silencio

Realización de la prueba el **14-Abril-2018 / 17:48h**

Ratio instantáneo músculo izquierdo  
 Ratio instantáneo músculo derecho

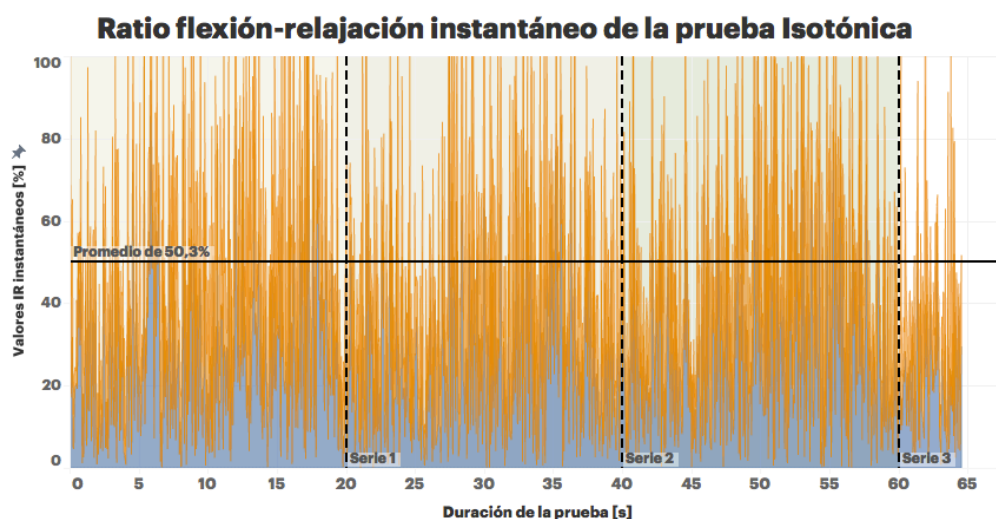


Figura 65. Descripción de valores de cada músculo y representación de la distribución de IR instantáneo.

Así pues, otra comparación interesante se posibilita al representar las señales EMG de los músculos de forma aislada y buscando la proporción con el *MVC máximo histórico*, tal y como se ha comentado en el apartado 4.5 de *análisis de datos*. De esta forma, fácilmente se puede comprender el margen que tiene cada músculo para trabajar bajo condiciones normales.

El gráfico de líneas de la figura 66 muestra esto y también corrobora la anterior, contemplándose un menor margen por parte del músculo izquierdo y el doble de carga de trabajo por su parte que por la del derecho.

Vemos, por tanto, que el conjunto de gráficos correspondientes a las figuras 65 y 66 son análogos y se complementan perfectamente entre sí, estableciendo un mismo tipo de gráfico desde dos perspectivas distintas.

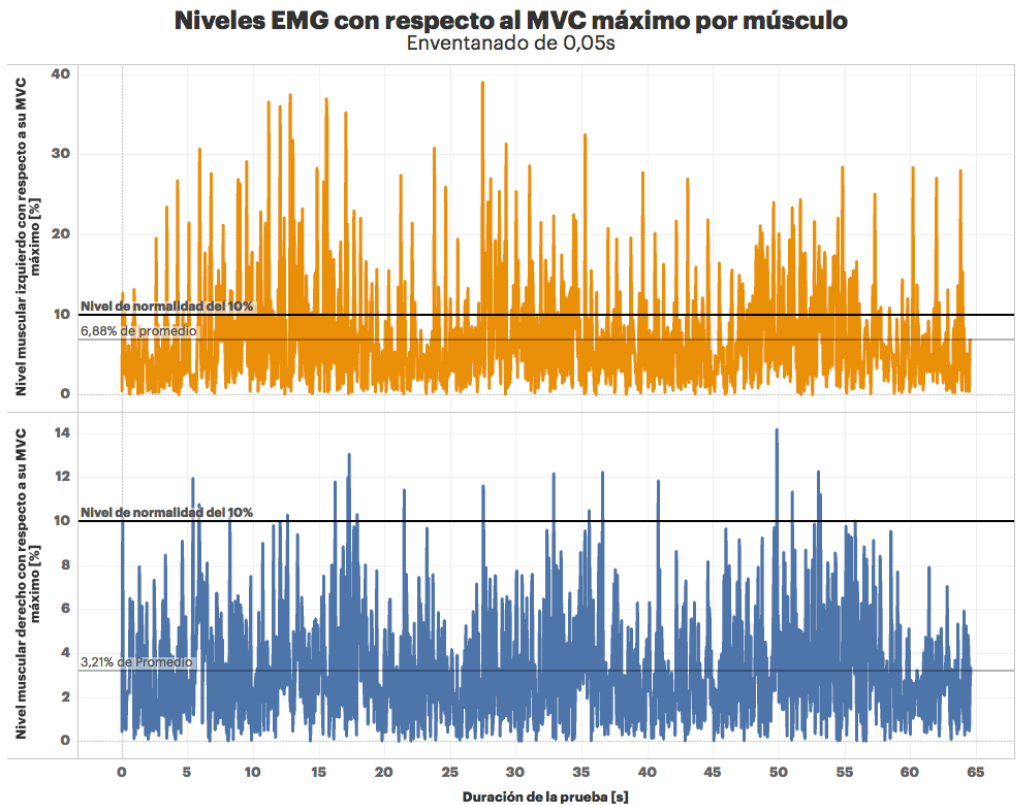
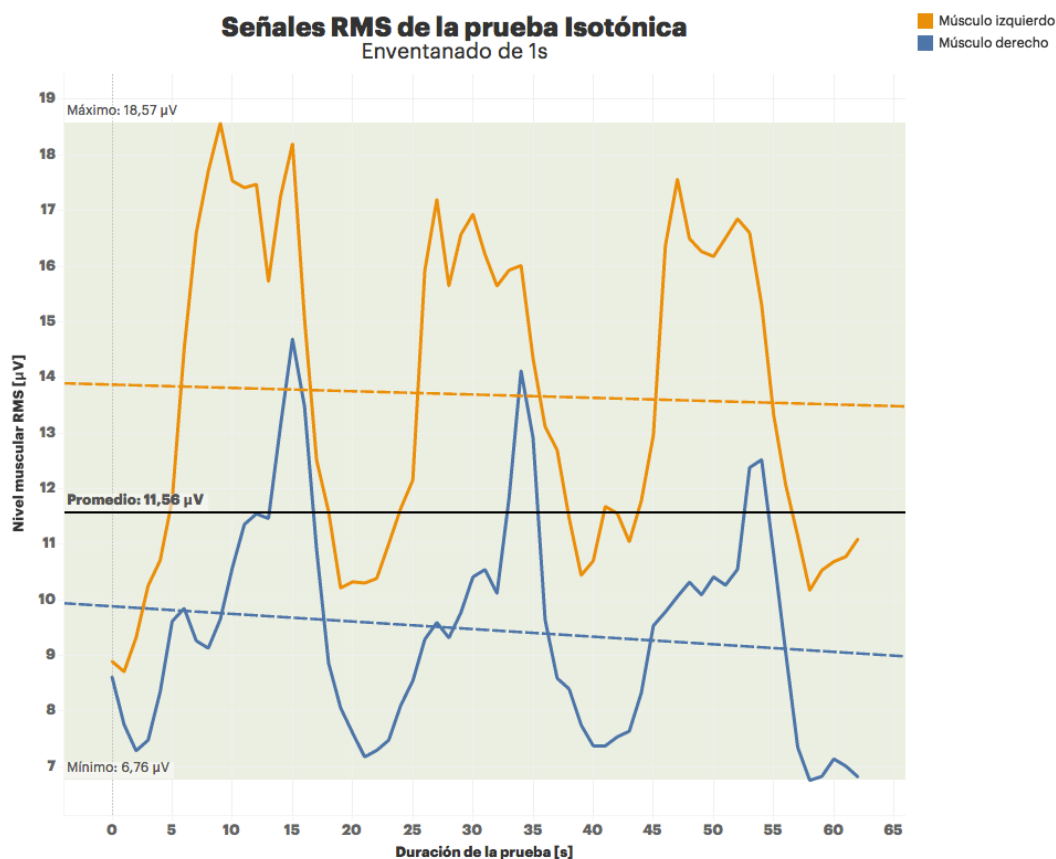


Figura 66. Descripción de valores de cada músculo y representación de la distribución de IR instantáneo.

Pasando a las señales de tipo RMS, se representa mediante la siguiente figura un modelo estadístico donde se aprecian los picos máximos y mínimos de cada una de las señales, así como el nivel promedio de la distribución. Se puede notar en este tipo de representación un mayor nivel de suavizado en las señales, y es que, tal y como se ha comentado previamente, en este punto se pasa de tener un tamaño de ventana de 0,05 segundos propio de las señales EMG, a uno de 1 segundo que caracterizan las señales RMS.

Siguiendo con los gráficos de líneas, el suavizado aquí se aprecia en la continuidad de la propia línea, y es que al fin y al cabo existen muchos menos puntos a representar. Este concepto es de gran ayuda en la percepción del gráfico, ya que los representados anteriormente requieren de un mayor nivel de atención, mientras que el gráfico RMS categorizado de la misma forma es mucho más fácil de asimilar.



**Figura 67. Representación de las señales RMS de la prueba isotónica con modelo de tendencia lineal.**

Este suavizado de señal permite, también permite mostrar con mayor nitidez y facilidad gráfica ciertos elementos estadísticos como por ejemplo las líneas de tendencia observadas en la figura de arriba para cada uno de los músculos.

Por otro lado, una forma alternativa y también de utilidad para visualizar el estado de los músculos es mediante el uso del que hasta ahora acostumbramos como eje vertical representado en modo horizontal, para así contemplar el tamaño de muestras desde la perspectiva de la simetría de los mismos. Teniendo en cuenta que estamos tratando con músculos izquierdo y derecho, una buena opción podría ser la siguiente:

### Señales RMS de la prueba Isotónica Enventanado de 1s

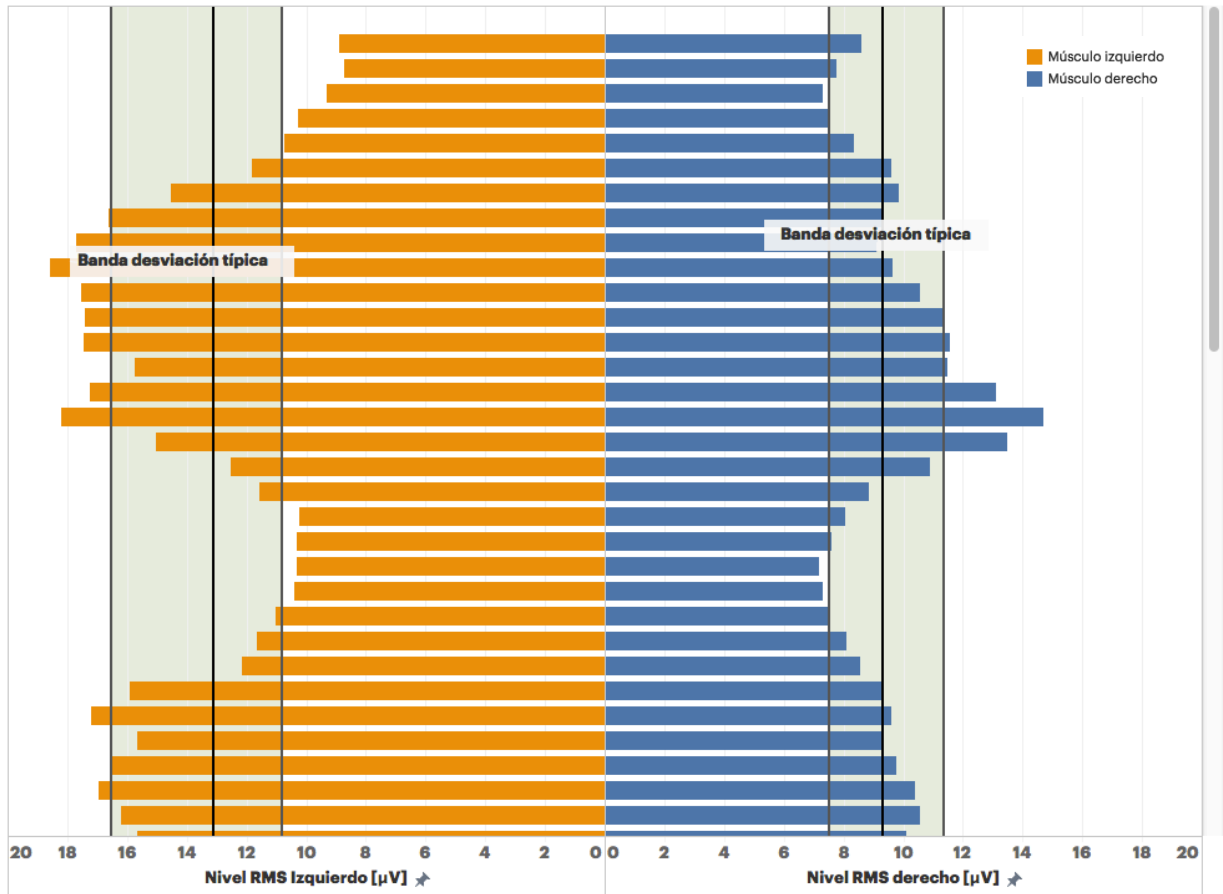


Figura 68. Histograma horizontal que representa la simetría de muscular en la prueba isotónica del paciente.

Estableciendo el mismo orden de eje horizontal para los dos músculos se puede contemplar en este histograma que la distribución no es ni mucho menos simétrica, sino que tiende a ser trabajada mayoritariamente por la parte izquierda. Esto también se corrobora mediante la banda de desviación típica, que al ser de un rango más amplio indica que trata valores más dispares con respecto a la mediana.

Esto se justifica debido a que los histogramas, al cambiar los ejes, pueden servir de comparación y mostrar simetría en lugar de evolución temporal. El ser humano tiende a concebir el tiempo como eje horizontal, mientras que el vertical es más intuitivo asociarlo a un valor útil, y en este caso, al estar el eje vertical centrado, la visualización es un balance.

Como última opción, se propone el uso de un diagrama de cajas para comprobar la *distribución de trabajo de los músculos* durante una prueba, y esto tiene la misma prueba de concepto que lo anterior: prescindir del eje temporal y centrarse en los valores útiles de la distribución.

## Rango de trabajo de los músculos en la prueba Isotónica

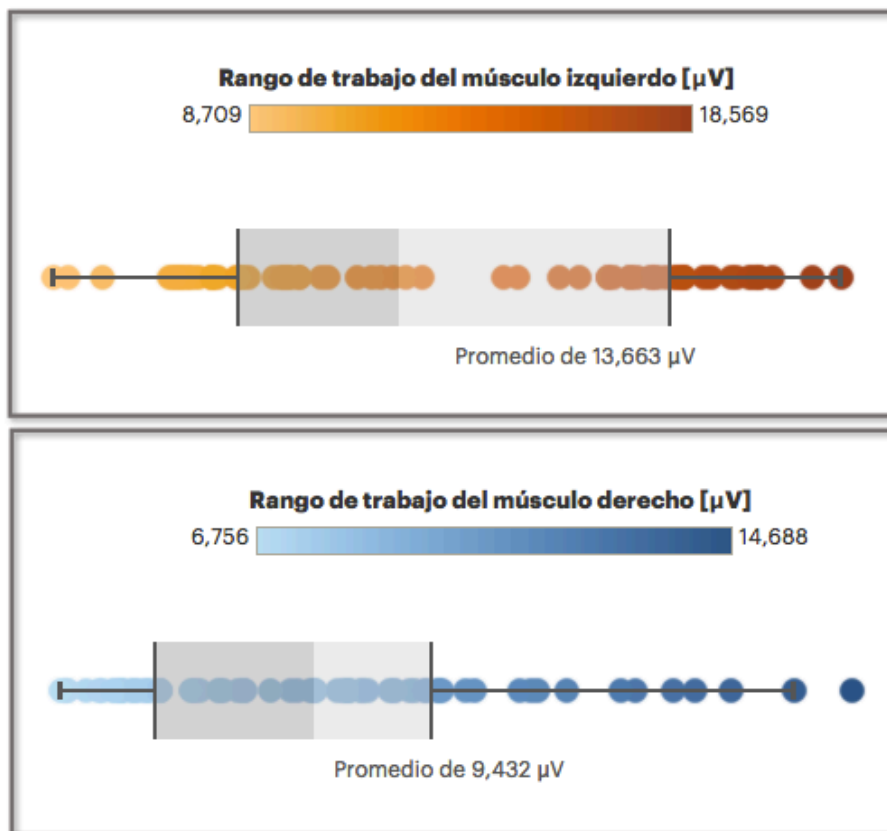


Figura 69. Representación de diagramas de cajas para observar el rango de trabajo de los músculos.

### 4.8.3 Comparación entre dos pacientes

Una vez visto el anterior repertorio de informes, se puede llevar a cabo una propuesta muy interesante en este ámbito de la Medicina del Deporte: la comparación entre casos de pacientes para la obtención de patrones estadísticos que sirvan de precedente para la exploración de patologías y detección de singularidades.

En este caso, al igual que el conjunto de etiquetas de datos es una combinación de las tratadas en los dos anteriores subapartados, las reglas de similitud que conforman la caracterización de gráficos también lo son. De cualquier modo, para este subapartado cabe destacar que:

1. La comparación entre los dos pacientes se indica mediante su *id*, que se puede reflejar mediante etiqueta de texto para así aprovechar el resto de propiedades del gráfico y mostrar más dimensiones y/o medidas.
2. Para no tener un elevado número de formas se representa el promedio de los dos músculos de cada uno de los dos pacientes – izquierda y derecha – con tal de no crear confusión con demasiados puntos y conseguir una comunicación clara.

Se puede comenzar dando paso a este tipo de herramientas visuales mediante la descripción básica de cada uno de los pacientes, comparando etiquetas semánticas

básicas y también otras sintácticas que hagan permitan obtener una visión general de ante qué tipo de paciente se encuentra el experto.

Es cierto que en este caso se toman como muestra dos pacientes contextualmente bien diferentes para ver comprobar la utilidad de las representaciones visuales en función del contexto y las características sintácticas que proporcionan las pruebas EMG, pero realmente no tendría por qué ser así. Una forma “objetiva” de saber ante qué tipo de paciente nos encontramos es mediante el RMS promedio dividido por el número de segundos de la prueba en cuestión.

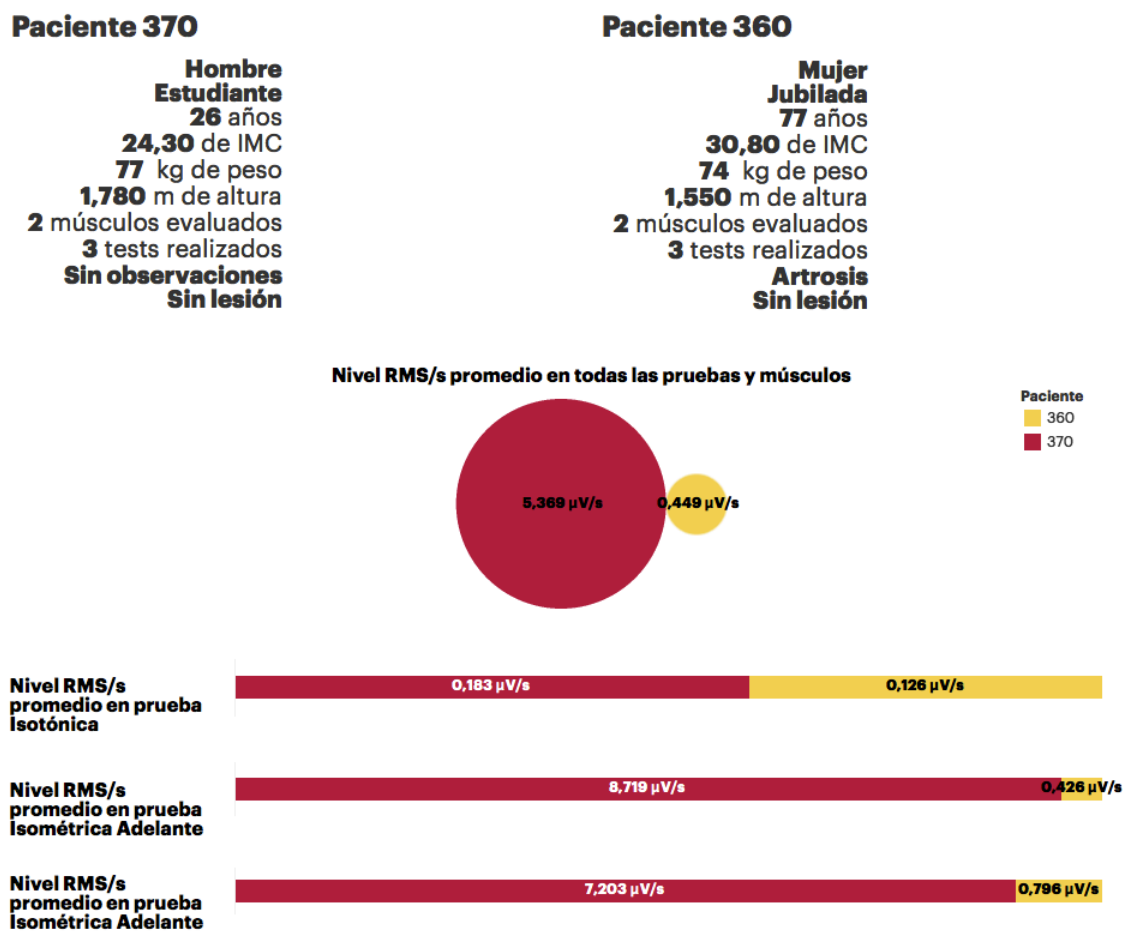


Figura 70. Descripción básica de los dos pacientes y sus niveles RMS/s promedio.

Para entrar más en detalle, se puede asociar la etiqueta semántica del *nivel de dolor* al nivel *RMS promedio* – en el eje vertical como parámetro cuantitativo principal – y también a otro parámetro sintáctico muy importante, el *MVC promedio* de cada una de las pruebas. Aquí, al desglosar en el eje horizontal por prueba, se necesitan puntos dispersos que caractericen mediante la forma el nivel de dolor, y mediante un degradado de colores el rango de niveles MVC como un parámetro cuantitativo asociado.

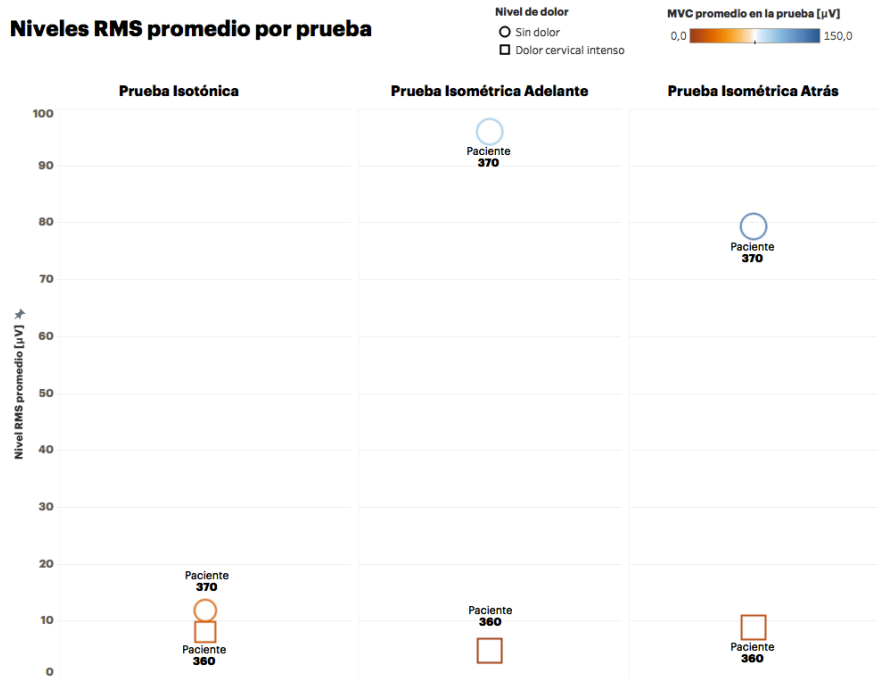


Figura 71. Representación de los niveles RMS promedios asociado al nivel de dolor y MVC de cada prueba.

Algo prácticamente igual ocurre si en lugar de disponer del nivel de RMS promedio, se tiene el nivel *MVC máximo* como valor cuantitativo primario, y se complementa en este caso con el *FRR promedio* como etiqueta asociada. Desde el punto de vista analítico, aquí se permite al experto comparar entre el máximo nivel de contracción (MVC) y la capacidad de extensión (FRR) del paciente en la prueba, por lo que solamente la uniformidad de la posición de los puntos ya debe dar una idea clara del estado de fatiga de los músculos.

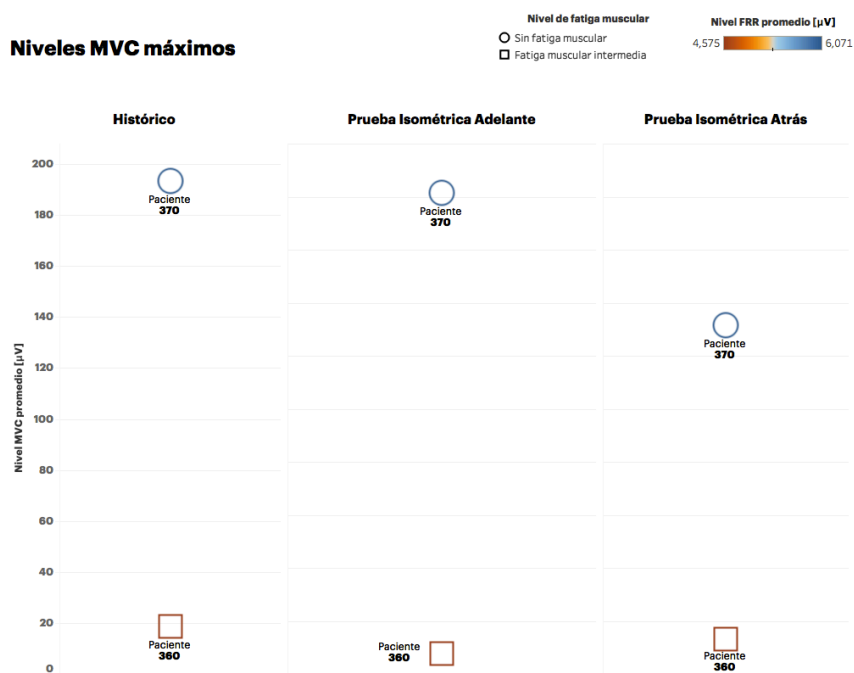


Figura 72. Representación de los niveles MVC máximos asociado al nivel de fatiga muscular y FRR.

Por último, otro aspecto a representar es la combinación de ángulos *onset* y *offset* y la estrecha relación que guardan los *tiempos de cambio* de estos ángulos con el *periodo de silencio*.

La combinación se plantea de forma que estos valores sintácticos se puedan relacionar con la etiqueta semántica de *condición física*, por lo que a ésta se le asigna el atributo de color, modo que:

1. Por un lado, se hace uso de un gráfico de puntos donde se disponen de los ángulos onset y offset en ejes cartesianos paralelamente a los tiempos de cambio asociados.
2. Por otro lado, se dedica un gráfico de dos burbujas que, mediante el tamaño de las mismas, indica el periodo de silencio promedio de los dos músculos sobre cada uno de los pacientes.

### Niveles Onset y Offset

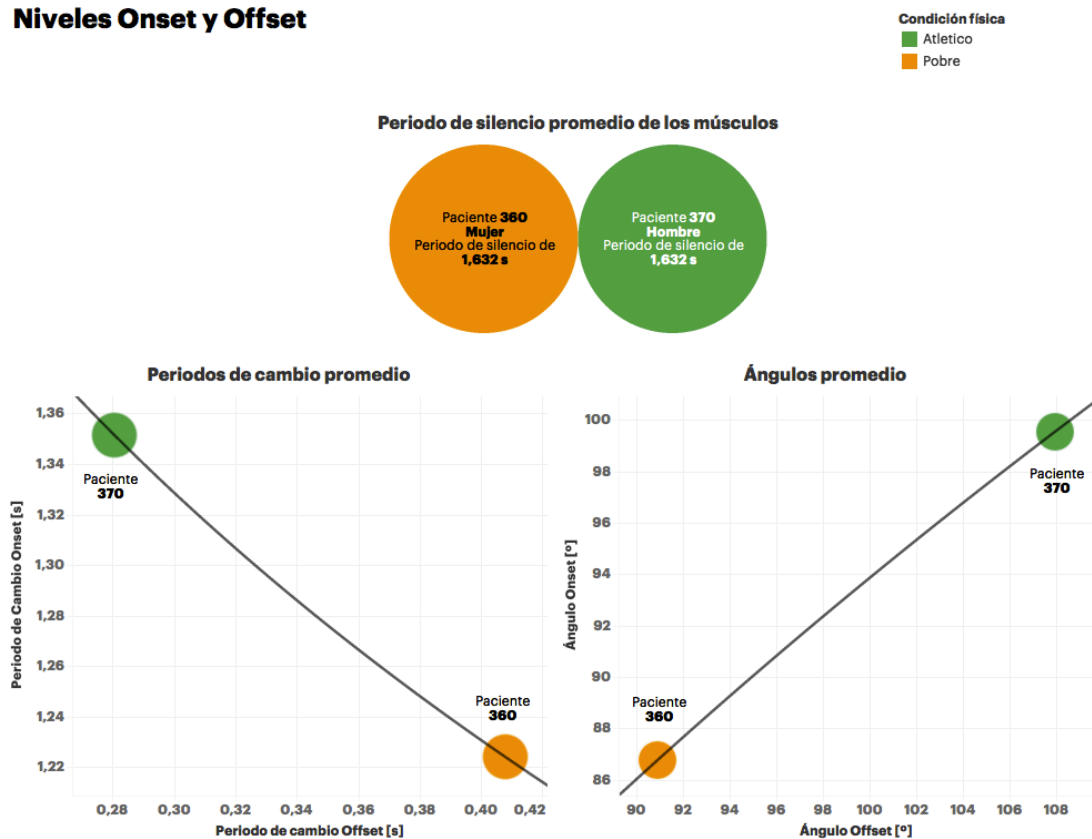


Figura 73. Representación de niveles onset y offset promedio de los músculos, así como el periodo de silencio.

#### 4.8.4 Comparación entre un paciente y el resto de población

Otra propuesta muy interesante en nuestro contexto de aplicación es la comparación de un caso aislado de paciente con los niveles promedios de población en su totalidad, para así tener un mayor nivel de abstracción analítico y obtener explicaciones sobre determinados niveles sintácticos o plantear hipótesis sobre segmentos de población en función de ciertas características semánticas.

En este caso, las reglas de similitud que conforman la caracterización de gráficos correspondientes a este tipo de informes también son una combinación de lo anterior. Destacar que:

1. En este caso la etiqueta de *id* no funciona de la misma forma que en el caso anterior. Ahora no equivale a dos valores aislados, sino que son todos los identificadores de la población, por lo que en la visualización se ha filtra cada etiqueta por el id del paciente seleccionado y luego, por otro lado, se toman los valores promedios de la población.
2. Aquí se combinar en mayor medida la cantidad de etiquetas sintácticas y semánticas, ya que es un caso de gráficos mixto que se puede explotar generando explicaciones a modo bidireccional. Esto es, por un lado, desde la perspectiva en la que el paciente puede destacar en etiquetas que se encuentran dentro de unos límites casi preestablecidos en la población (como por ejemplo en altura y/o peso); y por otro, desde el punto de vista en el que ciertas etiquetas del paciente (como la ocupación) pueden resultar determinantes para el resto de la población.

El primer conjunto de gráficos generado se corresponde de la misma forma al anterior, con la peculiaridad de que en las etiquetas de texto de la población se añaden porcentajes con los niveles de alteración y lesión.

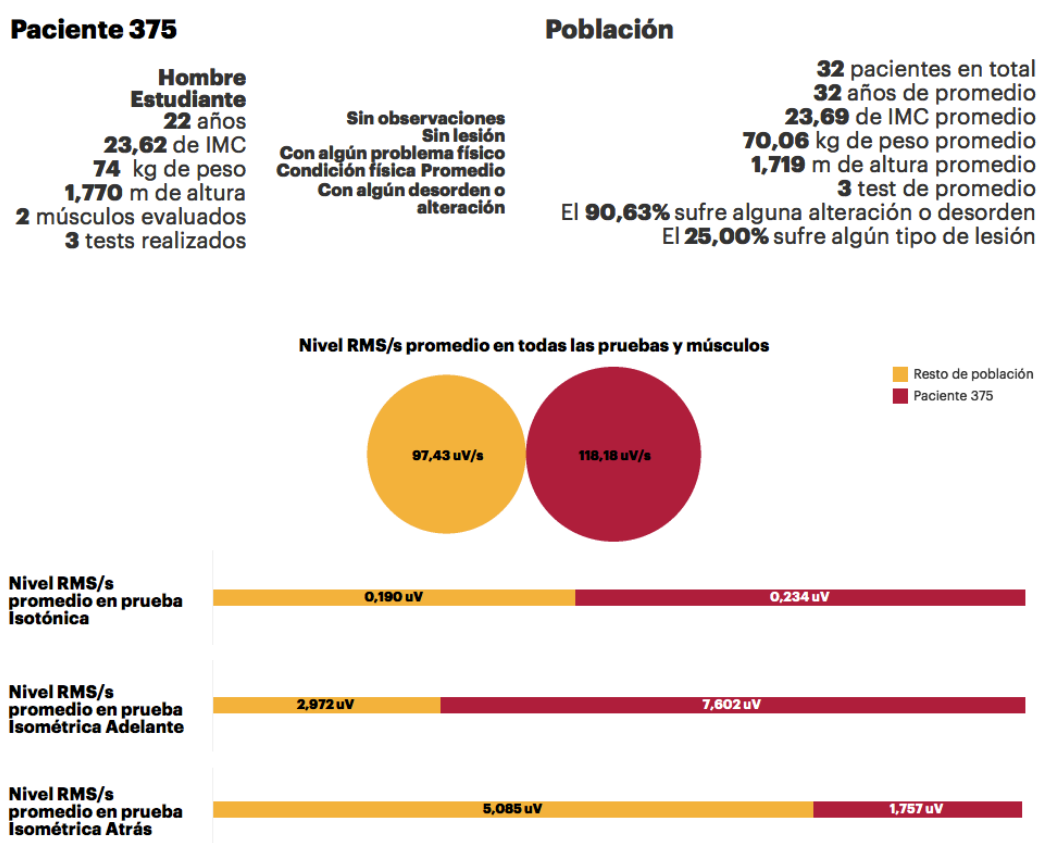


Figura 74. Descripción básica de un paciente en comparación con la población.

Por otro lado, para representar la singularidad de los *niveles históricos de MVC* del paciente con respecto al resto de población, se puede utilizar un diagrama de caja y bigotes combinado con puntos en los que se caracterice a dicho paciente con el atributo

del color, de la misma forma que antes. Así pues, este gráfico también permite contemplar si el nivel promedio del paciente se encuentra dentro de los niveles normales del modelo de distribución que dibuja el resto de población.

Además, para desglosar estos niveles MVC a modo de comparativa, el diagrama de caja se complementa con un gráfico de barras en el que se recojan las tres pruebas realizadas, en el eje horizontal, y sus niveles MVC promedio de los dos músculos, en el eje vertical.

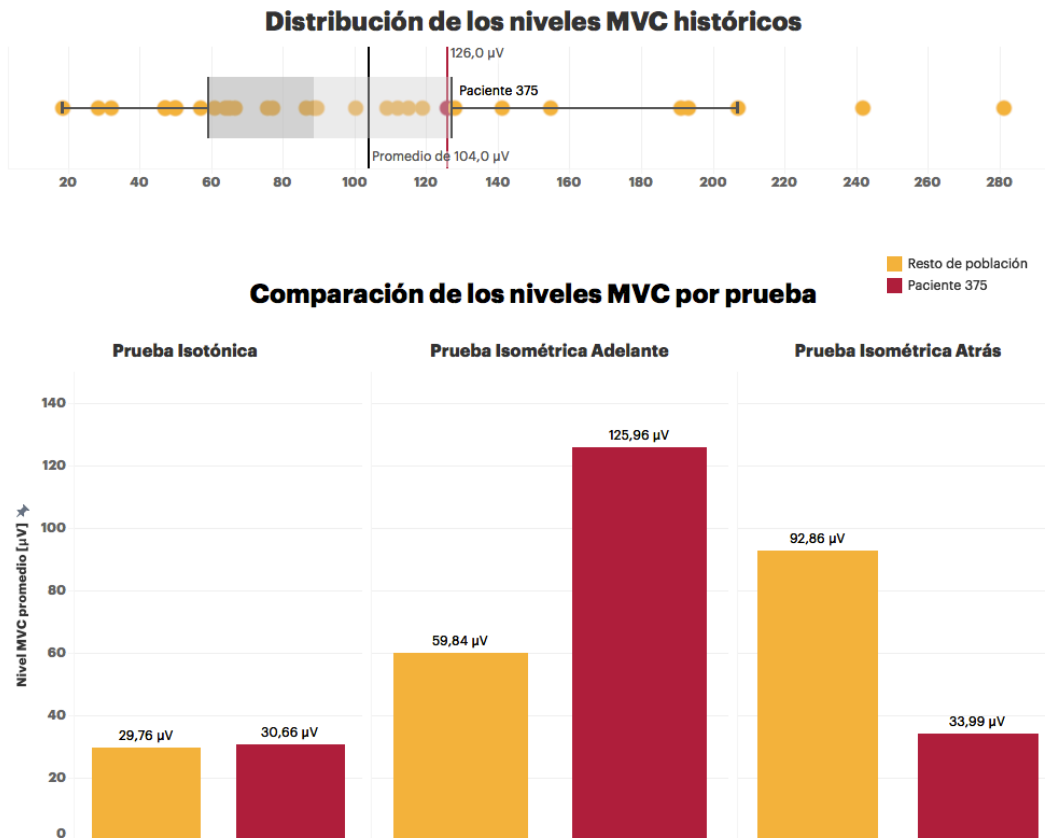
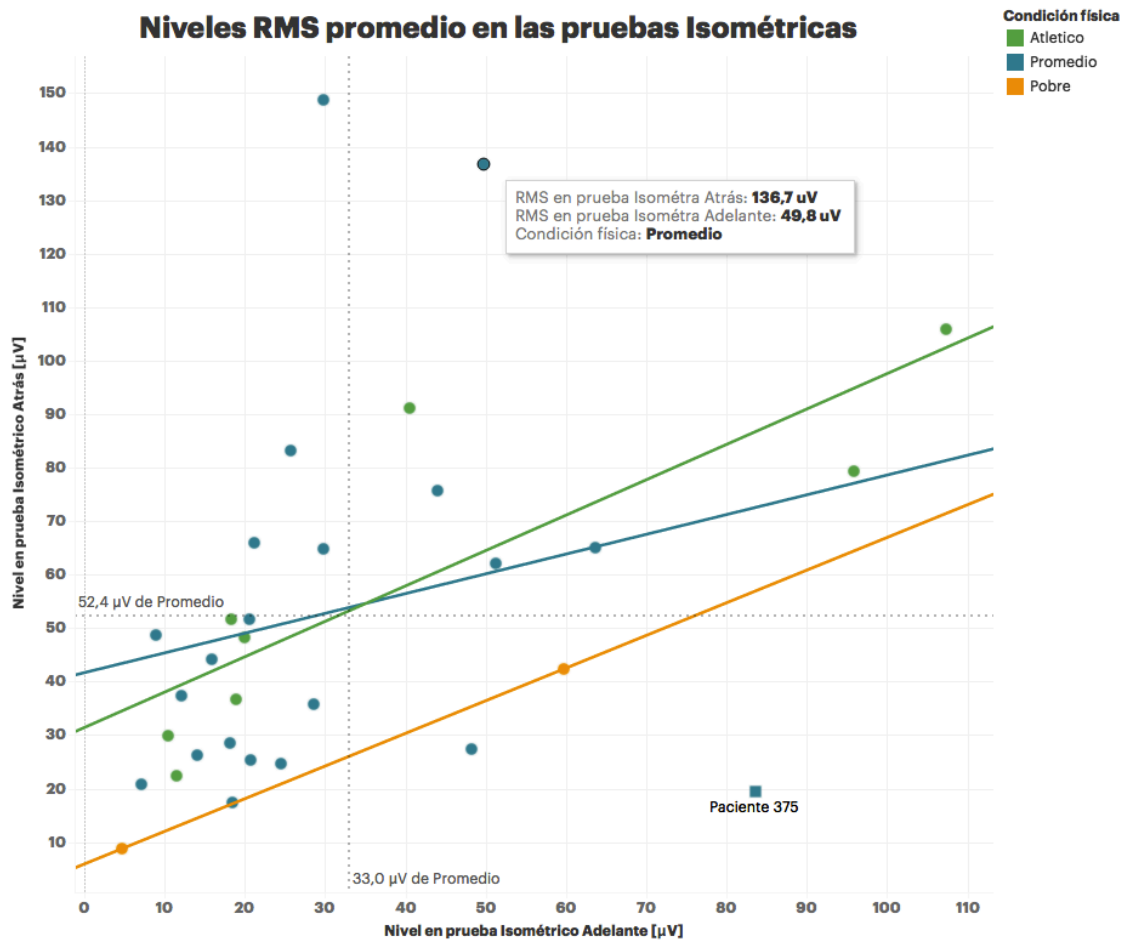


Figura 75. Representación de los niveles de MVC entre el paciente y la población.

Una vez visto esto, se pueden mapear los niveles de las pruebas isométricas, donde requiere un elevado esfuerzo por parte de las cervicales, con la condición física del paciente. Aquí se plantea la combinación de las dos pruebas isométricas en cada uno de los ejes (de forma indiferente, pero siempre siguiendo el mismo patrón) para representar toda la población de pacientes en la rejilla.

De este modo, con el atributo de color se puede distinguir la condición física de cada uno, complementando líneas de tendencia para cada uno de los segmentos de población. Aquí, una forma de singularizar al paciente es con la combinación de etiqueta de texto y cambio de forma; pues todos los casos están representados mediante objetos con forma de circunferencia mientras que el paciente a destacar tiene asociada forma cuadrada.



**Figura 76. Representación de los niveles RMS promedio en pruebas isométricas para asociarlo a la condición física de los pacientes.**

Por último, se representa un diagrama de burbujas en el que se pretende contextualizar la capacidad de extensión promedio de los dos músculos de los pacientes en la población, y destacar mediante etiqueta de texto el paciente en cuestión. Para ello, entra en juego la etiquetas de *FRR* y se decide complementar con la *profesión* de los pacientes y sus posibles *problemas ergonómicos*.

En un diagrama de burbujas, tal y como ya se ha dicho, se caracteriza el tamaño de las burbujas mediante el número de pacientes. La etiqueta cuantitativa es la correspondiente al *FRR*, por lo que se caracteriza con el atributo de color mediante un rango en tonos graduales. Por su parte, la etiqueta de problema ergonómicos se puede caracterizar por columnas ya que se define siempre mediante dos valores, o afirmativo, o negativo. Por último, la etiqueta de profesión es un atributo de texto en cada una de las burbujas.

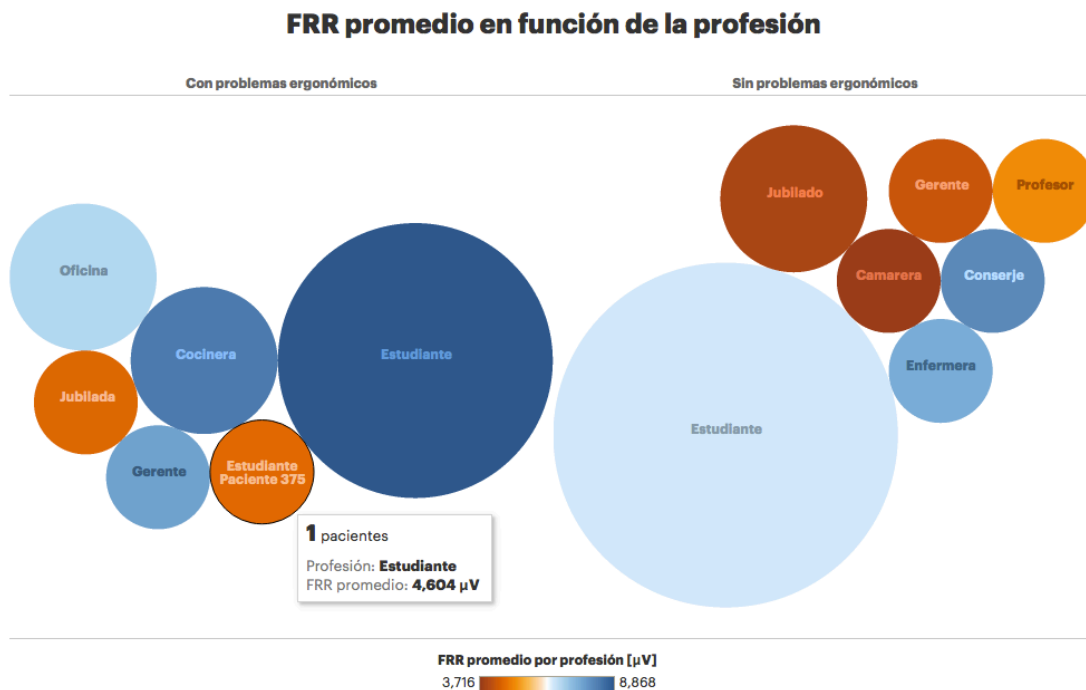


Figura 77. Representación de la capacidad de extensión de los músculos cervicales por profesión y ergonomía.

#### 4.8.5 Ejemplo pormenorizado de utilización de los informes

Una vez estudiadas las formas de visualización susceptibles a ser reutilizadas y adaptadas según el modelo de informe requerido, se propone a continuación un ejemplo sencillo que permite comprobar el verdadero funcionamiento del mecanismo implementado. En él, el experto médico realiza una solicitud al sistema y recupera el informe correspondiente unos determinados huecos a completar con los gráficos guardados en la base de casos.

Para simplificar y seguir una misma línea de ejemplo, se propone el mismo caso que el comentado al final del subapartado 4.7.2, es decir, un informe de inspección de un paciente de manera individual. De este modo, se procede a realizar dos peticiones de pacientes distintos para comprobar que mediante el mismo procedimiento de consulta se pueden obtener gráficos distintos, adaptados a cada uno de los pacientes en función de las características propias y atendiendo a determinadas reglas de similitud.

Antes de nada, observemos el lenguaje de etiquetas que permite la comunicación entre el fisioterapeuta con el sistema para la reutilización de gráficos, y que por tanto forma parte del sub-ciclo CBR:

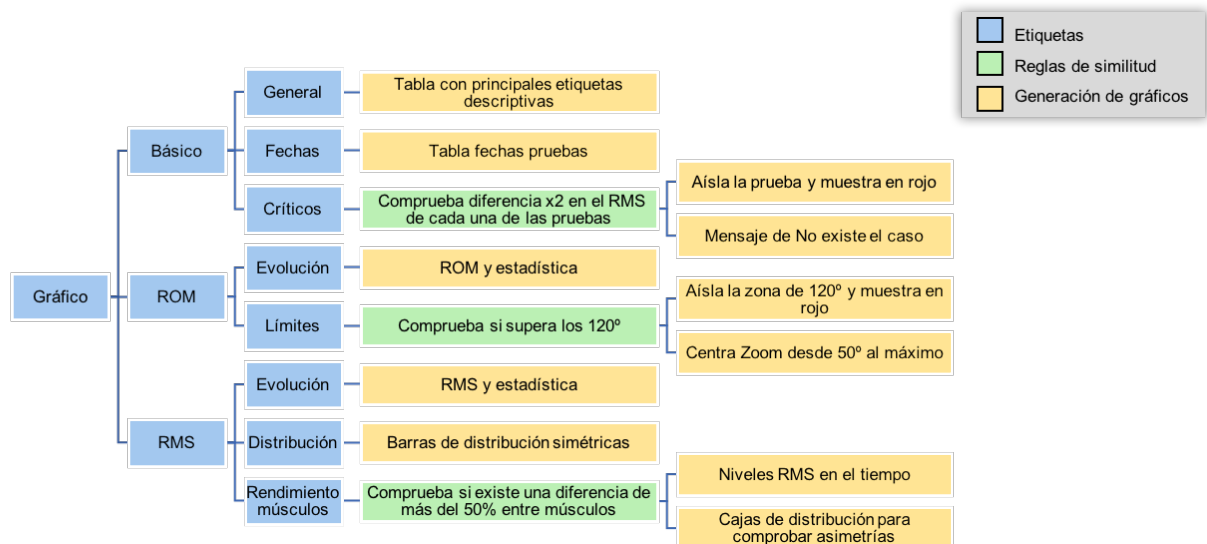


Figura 78. Etiquetas de lenguaje fisioterapeuta-sistema para la reutilización y adaptación de gráficos.

Supongamos que el fisioterapeuta desea obtener un gráfico sobre el paciente con ID 360, por lo que primeramente realiza la petición que consigue suplantar lo que sería la recuperación de una plantilla de informe.

Hasta este punto se tienen, por tanto, las etiquetas de *Informe*, *Inspección*, *Paciente*, 360.

```

----- ACTUALIZAR LAS FUENTES DE DATOS DE INFORMES -----
1. Inspección de población
2. Inspección de paciente
3. Comparación entre dos pacientes
4. Comparación entre paciente y población
5. Exit
-----
Enter your choice [1-5]: 2
Introduce id de paciente: 360

```

Figura 79. Etiquetas: *Informe*, *Inspección*, *Paciente*, 360.

Una vez se actualizan todas las fuentes de datos con la caracterización adecuada, el sistema ya conoce los gráficos candidatos a ser reutilizados, pues son aquellos que coincidan con las etiquetas que se tiene hasta ahora. De manera seguida, lo que se pretende recuperar es el gráfico que se adapta a los huecos de este tipo de informe para este único paciente. Es por ello que el fisioterapeuta, que en mayor o menor medida conoce el funcionamiento del sistema, ha de decidir si recuperar un *gráfico de análisis básico*, uno que defina el *rango de movimiento (ROM)* del paciente, u otro que se encargue de definir los *niveles RMS* que el paciente ha ejercido en la prueba.

Pues bien, supongamos que elige la tercera de las opciones, la del análisis RMS dentro de la prueba EMG (en este caso isotónica, por simplificar el proceso). Esta etapa del proceso supone un punto de inflexión para el fisioterapeuta, ya que dispone de tres opciones para continuar con el análisis.

1. La primera de ellas se corresponde con un análisis en el que se recupera automáticamente un gráfico de líneas que muestra la evolución temporal de la

- prueba RMS para cada uno de los músculos – izquierda y derecha – con las etiquetas estadísticas que componen dicho análisis.
2. La segunda forma de análisis se corresponde con un gráfico de barras horizontales (histograma) que muestra los niveles de ambos músculos – izquierda y derecha – de modo que represente la distribución de cada uno de ellos en un vistazo general.
  3. Por último, la tercera opción permite al fisioterapeuta “confiar” en cierto modo en el sistema y otorgarle la capacidad de decisión en el gráfico a elegir para comprobar el rendimiento de los músculos. Esta decisión puede ser útil a la hora de un experto médico elegir un gráfico para posteriormente realizar análisis más profundos en función de la situación de equilibrio/desequilibrio de los músculos izquierdo y derecho. Es decir, a pesar de que no es un indicador para conocer de primera mano si el paciente se encuentra sano o enfermo, esta elección tiene la firmeza suficiente como para permitir al fisioterapeuta indagar en unos aspectos u otros.

De esta forma, se elige para el paciente 360 la visualización RMS que permite al sistema elegir entre un gráfico u otro en función de una regla de similitud.

Hasta este punto se tienen, por tanto, las etiquetas de *Informe, Inspección, Paciente, 360, Gráfico, RMS, Verificación.*

```

----- GENERACIÓN DE GRÁFICOS PARA INSPECCION DEL PACIENTE 360 -----
1. Descripción básica
2. Range of Movement
3. Flexion-Relaxation Ratio
4. Niveles RMS
-----
Inspección, Paciente, 360, RMS
----- GRÁFICO RMS PARA PACIENTE 360 -----
1. Evolución temporal
2. Distribución de músculos
3. Verificación del rendimiento
-----
Enter your choice [1-5]: 2
Introduce id de paciente: 360
Enter your choice [1-4]: 4
Enter your choice [1-3]: 

```

**Figura 80.** Etiquetas: *Informe, Inspección, Paciente, 360, Gráfico, RMS, Verificación.*

En este punto entran en juego las reglas de similitud establecidas. Si la diferencia porcentual entre los músculos del paciente 360 supera el 50%, estamos hablando de que posee un desequilibrio elevado, mientras que si no lo supera nos encontramos ante un paciente cuyos niveles RMS son fácilmente representables y comparables en una línea de tendencia temporal, tal y como se muestra en la siguiente figura:

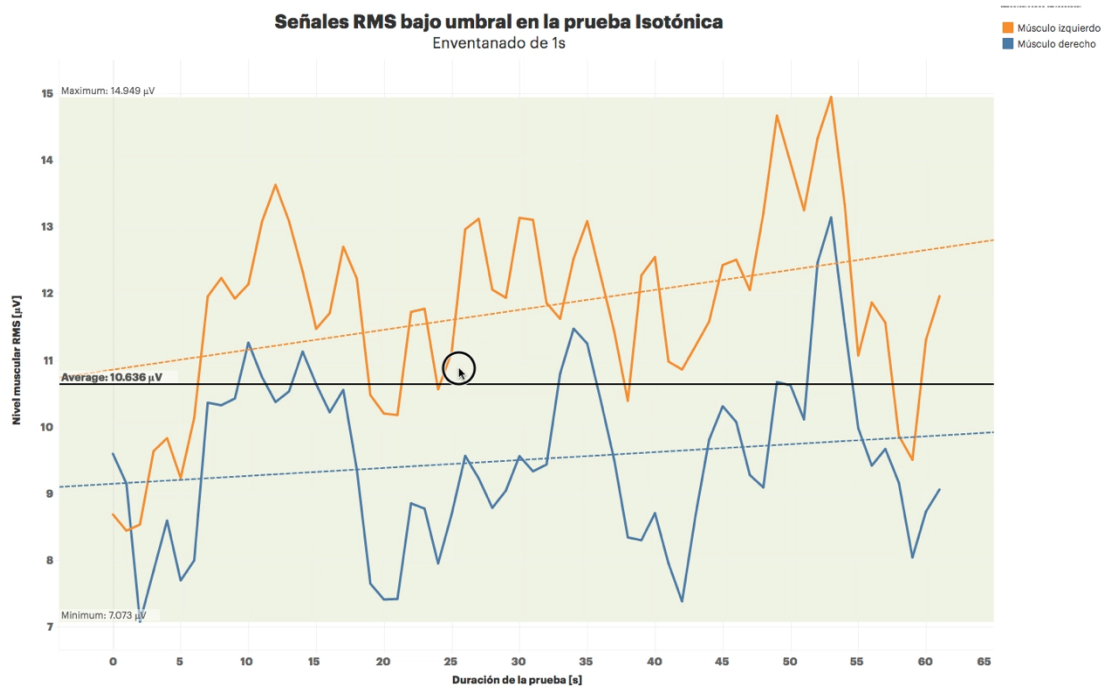


Figura 81. Representación del gráfico recuperado para el paciente 360.

Se puede observar que ambos músculos se encuentran dentro del mismo promedio de valores, y que las líneas de regresión poseen una pendiente muy similar. Es por ello que se puede decir que las señales RMS del paciente 360 se encuentran dentro del umbral de “normalidad”.

In, Supongamos que el fisioterapeuta desea obtener un gráfico sobre el paciente con ID 363, por lo que de nuevo realiza la petición que consigue suplantar lo que sería la recuperación de una plantilla de informe para este otro paciente.

```

----- ACTUALIZAR LAS FUENTES DE DATOS DE INFORMES -----
1. Inspección de población
2. Inspección de paciente
3. Comparación entre dos pacientes
4. Comparación entre paciente y población
5. Exit
-----
Enter your choice [1-5]: 2
Introduce id de paciente: 363
  
```

Figura 82. Etiquetas: *Informe, Inspección, Paciente, 363*.

Seguidamente, se vuelven a actualizar todas las fuentes de datos con la caracterización adecuada para volver a conocer los gráficos candidatos. Suponemos el mismo caso que antes: la recuperación de un gráfico RMS en el que el sistema nos ayude a conocer a prior el estado de equilibrio muscular del paciente. Esto implica que el fisioterapeuta siga el mismo procedimiento que el indicado anteriormente y se disponga de las etiquetas siguientes: *Informe, Inspección, Paciente, 363, Gráfico, RMS, Verificación*.

La esencia de este pequeño ejemplo es que para el paciente 363 se cumpla el caso contrario al establecido en la regla de similitud para el paciente 360, es decir, que tenga un nivel de desequilibrio muscular mayor a 50 puntos porcentuales.

Contemplamos la decisión gráfica que devuelve el sistema:

## Desequilibrio de los músculos en la prueba Isotónica

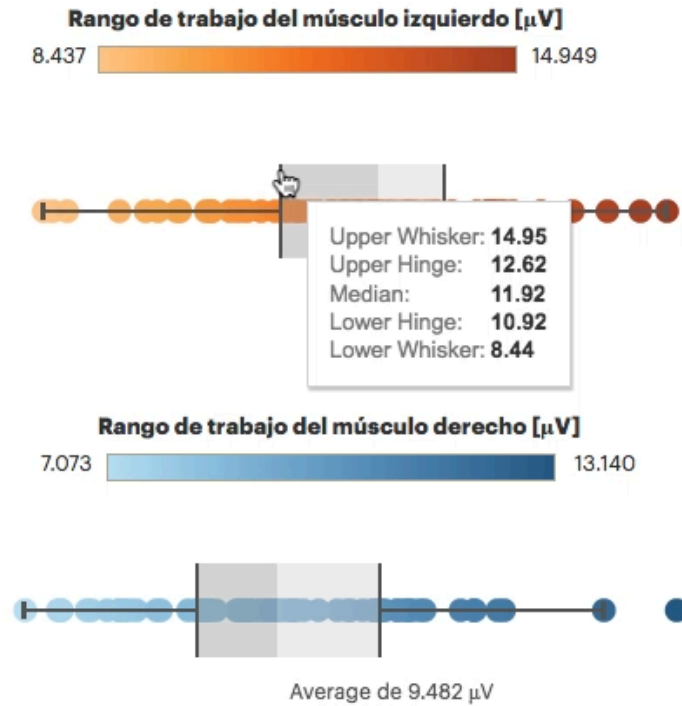


Figura 83. Representación del gráfico recuperado para el paciente 363.

Se observa que en este caso el gráfico elegido representa dos diagramas de cajas en los que se puede ver la completa discrepancia en cuanto a niveles RMS de los músculos izquierdo y derecho. Por una parte, la distribución de niveles del músculo izquierdo se encuentra en todo momento por encima del promedio de la distribución del músculo derecho; y por otro lado, la distribución de niveles del músculo derecho recae por debajo del promedio de la distribución del músculo izquierdo.

Además, este análisis de desequilibrio muscular puede resultar útil incluso para detectar posibles fallos en la ejecución de las pruebas EMG, ya que en caso de que existieran demasiados valores atípicos (fuera de los bigotes), podría catalogarse como una prueba fallida.

## Capítulo 5. Legalidad en el contexto

El estudio práctico que conforma este Trabajo Final de Máster se basa en un caso de población real, y por tanto, para su desarrollo se toman conjuntos de datos totalmente verídicos. Tal y como se ha visto en el anterior capítulo, los datos caracterizados y representados visualmente son de carácter personal y, como fruto de ello, altamente valiosos. Además, en este tipo de procedimientos es un hábito muy común el hecho de que sean objeto de manejo mediante diversas plataformas tecnológicas, por lo que también se puede decir que tienen un gran alcance y son susceptibles a ser tratados en situaciones muy dispares.

Es por ello que la gestión de los datos recogidos tanto en este proyecto como en cualquier proceso de un sistema sanitario con actuación en el campo de la Medicina del Deporte se encuentra sujeto a normativas reguladoras impuestas tanto en España como por parte de la propia Unión Europea en lo referente a la protección de dichos datos.

Por este motivo, y sumándole que en el presente año la nueva regulación de protección de datos ha pasado a ser totalmente aplicable, mediante el presente capítulo se procede a repasar los fundamentos y procedimientos legales que implican el tratamiento y la exposición de este tipo de datos.

### 5.1 Anonimización de los datos

Antes de nada, cabe destacar que los datos tomados en este proyecto son gestionados íntegramente por mDurance, tanto desde la perspectiva del procesamiento como por la parte del almacenamiento y tratamiento de los mismos. Es por ello que los partícipes de este Trabajo Final de Máster han sido autorizados explícitamente por el Delegado de Protección de Datos (DPO) de mDurance para capturar datos únicamente de los sujetos que forman parte del experimento de investigación que abarca el presente Trabajo Final de Máster, pero en ningún momento de ningún otro paciente en relación con mDurance o cualquiera de las clínicas que cuentan con el servicio de su plataforma digital.

### 5.2 Punto de partida

Los conjuntos de datos que se ponen sobre la mesa en este trabajo son realmente de carácter personal, ya que contienen información numérica, alfabética, y sobre todo gráfica, aplicando tanto a personas identificadas como identificables [57]. Esto implica que los partícipes de un experimento como el nuestro posean el **derecho fundamental a la protección de datos** [58], el cual establece que todo paciente posee la facultad de:

- Controlar sus datos personales
- Disponer de dichos datos.
- Decidir sobre todos estos datos.

A día de hoy el cumplimiento de este derecho fundamental en nuestro país tiene como tutela la Agencia Española de Protección de Datos (AEPD), que posee la autoridad de controlar el cumplimiento de toda la normativa sobre protección de datos personales. La base constitucional conforma el pilar fundamental sobre el que se sustenta [59]:

*“Art 18.4 de la Constitución Española: La ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos.”*

Esta base tiene aplicación mediante los reglamentos encargados de asegurar una serie de derechos fundamentales para los ciudadanos, los cuales son objeto de estudio en subapartados posteriores. No obstante, antes de pasar a un análisis de la regulación en mayor profundidad es recomendable tener clara la diferenciación de roles que aplican a nuestro experimento en materia legislativa:

<b>Rol</b>	<b>Definición</b>	<b>Correspondencia</b>
Interesado/Afectado	Persona física a la cual pertenecen los datos recogidos en la prueba médica	Paciente
Responsable de los recursos	Entidad encargada de la recolecta y la formalización de los datos	mDurance
Encargado del tratamiento de los datos	Persona comprometida tanto al almacenamiento y mantenimiento de los datos como al seguimiento de buenas prácticas en todos los procedimientos.	Autor de este Trabajo Final de Máster
Usuario	Persona autorizada para acceder al historial clínico y otros recursos	Autor de este Trabajo Final de Máster
Responsable de la seguridad de los recursos	Entidad a la que el interesado asigna consentimiento formalmente desde un primer momento al requerir uno de sus servicios	mDurance

**Tabla 11. Correspondencia de roles en el presente TFM desde el punto de vista legislativo.**

### **5.3 Antecedentes reguladores: LOPD**

La regulación de protección de datos que aplicaba hace unos meses en día se enmarca según dos bases normativas [60]:

1. “Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de carácter Personal (LOPD)”.
2. “Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la LOPD”.

Básicamente, la LOPD implica cumplir una serie de premisas. No es objeto de este trabajo tratarlas en profundidad, pero sí definir las para valorar el importante rol del responsable de datos y también ver a posteriori la evolución en los cambios legislativos [61].

Estas premisas se basan en:

- Inscripción de los ficheros en la AEPD.
- Nivel de consideración en los datos especialmente protegidos.
- Acceso a los datos por cuenta de terceros.
- Calidad de los datos en todo su proceso.
- Seguridad de los datos según niveles.

- Cumplimiento de los derechos ARCO: Acceso, Rectificación, Cancelación y Oposición.
- Deber de información al interesado.
- Deber de secreto en cualquier fase del tratamiento.
- Consentimiento tácito del afectado.
- Control total en la comunicación de los datos.
- Control especial en la transferencia internacional de los datos.

#### 5.4 Nueva regulación en vigor: GDPR

A partir del 25 de mayo de 2018 se convirtió en plenamente aplicable el Reglamento UE 2016/679 del Parlamento Europeo y del Consejo; o lo que es lo mismo, el Reglamento General de Protección de Datos (GDPR), de 27 de abril de 2016 [62], relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la LOPD.

En dicho Reglamento, se definen los datos correspondientes a la Medicina del Deporte como aquellos relativos a la salud física o mental de un individuo que proporcionen información acerca de su salud mediante cualquier prestación de servicios de atención sanitaria [63]. El GDPR, a diferencia de la LOPD, define este tipo de datos en los puntos 53 y 54 como **Especialmente Protegidos**, motivo por el cual se tengan que cumplir una serie de condiciones adicionales para su tratamiento.

Bien es cierto que en la aplicación del cambio legislativo algunas de las premisas básicas establecidas en la LOPD se mantienen: como los derechos ARCO, el deber de secreto y la inscripción de ficheros. No obstante, también se modifican e intensifican otros muchos aspectos que se han de tener en consideración. Se resumen a continuación para contemplar cómo afectan en nuestro caso de estudio.

##### 5.4.1 *Consentimiento explícito*

La primera novedad destacable es que, recurriendo al artículo 9 del GDPR [64], el consentimiento debe ser manifestado de manera por el interesado de manera explícita y recogida por escrito. Únicamente se podría encontrar implícito para proteger intereses vitales, tanto del interesado como asociados; y para fines de medicina preventiva, asistencia laboral e interés público [65].

Por la parte que nos influye, esta indispensable tarea se ha realizado con todos y cada uno de los pacientes de forma prioritaria, pues cada uno de ellos firmó consentimiento para realizar el experimento a través del soporte digital que mDurance ofrece en su aplicación móvil.

##### 5.4.2 *DPO*

Otra de las nuevas introducciones del GDPR es la figura del DPO [66]; encargado, según los artículos 37, 38 y 39, de prestar ayuda y soporte a los responsables de ficheros sobre el cumplimiento de la normativa. En este caso el DPO requiere de formación y competencias específicas con tal de conseguir que mDurance se adapte a la nueva legislación como proveedor de herramienta de salud digital, evitando así cualquier

conflicto de intereses entre otros grupos de trabajo, como bien son los involucrados en este Trabajo Final de Máster.

#### **5.4.3 Aumento del nivel de información**

Además, con el nuevo Reglamento los pacientes tienen el derecho de recibir una cantidad de información más elevada por parte de los responsables de sus datos [67], como por ejemplo:

- Identidad y datos del DPO.
- Base jurídica para el tratamiento.
- Plazos de conservación de la información.
- Existencia de decisiones automatizadas.
- Previsión de transferencia internacionalmente.
- Derecho a presentar una reclamación ante la Autoridad de Control.

En la realización del experimento, a todos y cada uno de los partícipes se les informó del objetivo académico que tiene el experimento, comunicando explícitamente que los datos personales son tomados con tal de ser caracterizados y representados tanto de forma individual como junto a otros de un mismo tipo de una población.

Esto se ha hecho de forma primordial, antes la ejecución de cualquier prueba EMG. Además, se presentó mDurance como entidad a cada uno de los pacientes, el tremendo potencial que posee su herramienta de salud digital, y la garantía de fiabilidad en la protección de sus datos.

#### **5.4.4 Medidas organizativas y de seguridad**

Por otro lado, con la nueva regulación ya no se estipulan medidas de seguridad por niveles, sino que aplican medidas en función del riesgo que puedan ocurrir en el tratamiento de datos. Es por ello que nuestros datos, al considerarse Especialmente Protegidos, son objeto de diseño en base a unas medidas organizativas y de seguridad conforme a al riesgo que aplique en cada nivel de vulnerabilidad.

Es por esto que, de forma general, perfiles sanitarios como mDurance se ven obligados a cumplir internamente con unas implicaciones técnicas basadas en unas estrategias determinadas con tal de evitar brechas de seguridad que originen la destrucción, pérdida o modificación de forma accidental de los datos personales de los deportistas. Esto no solamente aplica a los datos recogidos en este experimento sino sobre todos aquellos asociados a su arquitectura tecnológica de recogida, almacenamiento y tratamiento de datos.

Estas estrategias son [68]:

- **Minimizar** la cantidad de datos personales recogidos.
- **Agregar** procesos técnicos a los datos.
- **Ocultar** los datos y sus interrelaciones para que no queden visibles.
- **Separar** los datos personales de forma distribuida.
- **Informar** a los interesados para proporcionar transparencia.
- Permitir y dar soporte al interesado para que pueda **controlar** sus datos.
- **Cumplir** con la política de privacidad.

- **Demostrar** el cumplimiento del Reglamento.

Así pues, en la tabla siguiente se plasman estas estrategias enlazadas con cada una de las técnicas de implementación oportunas según el nivel en el que se encuentre el proceso de tratamiento de datos:

Capa	Estrategia	Implementación
Fuente de datos	Minimizar Minimizar Agregar Ocultar Ocultar Controlar Informar	Selección minuciosa de los datos Monitorización continua del nivel de riesgo Pseudoanonimización Herramientas de cifrado Herramientas Enmascaramiento Recabar consentimiento del afectado Transparencia para el afectado
Integración capas de servicio	Agregar Ocultar	Pseudoanonimización Herramientas de cifrado
Almacenamiento Cloud	Ocultar Ocultar Separar	Mecanismos de cifrado Autenticación y control de acceso Almacenamiento distribuido
Explotación de datos	Agregar	Pseudoanonimización

Tabla 12. Implementación tecnológica y equivalencias en el caso que ocupa.

Es importante destacar que la capa donde mayor número de técnicas hay que emplear es la primera, la relativa al origen de los datos, pues como ya se sabe los dispositivos médicos de sensado en el mundo de *Internet of Things* son notablemente heterogéneos en cuanto a firmware, sistema operativo empleado, protocolos de comunicación, parches de seguridad... [69]. Esto, y así como todos los puntos comentados en la tabla 12, claramente aplica a nuestro caso de estudio de EMG. De esta forma, en el transcurso del trabajo se aplican las medidas de protección en función de:

- **Protección desde el diseño y por defecto** (Artículo 25): En este proyecto se persigue el fin establecido desde un primer momento, representar gráficamente los datos de los pacientes, lo cual no implica usar los datos en cualquier otra medida. Así pues, se estipula la regulación del tratamiento de datos por cuenta del destinatario y cumpliendo las directrices reguladoras que también sigue mDurance.
- **Seguridad de tratamiento** (Artículo 32): Se conocen y contemplan en todo momento los riesgos de brechas de seguridad que existen en el tratamiento. Para asegurar esto, se trabaja en todo momento con plataformas tecnológicas sofisticadas que aplican técnicas de pseudoanonimización y cifrado; evitando en todo momento el manejo de datos desde fuentes desconocidas. Es por esto que en nuestro proyecto se puede garantizar la disponibilidad de los sistemas para cumplir la integridad y acceso a datos en el proceso de tratamiento.
- **Violación de la seguridad de los datos** (Artículos 33 y 34): Para evitar violar la intimidad de los pacientes, bajo ningún concepto se muestra algún tipo de enfermedad, dolor o alteración. Es más, las cantidades de datos se minimizan siempre que se puede, evitando en todo momento representar nombres o identificadores de correo electrónico.

- **Evaluación del impacto** (Artículo 35, merece ser comentado con más detenimiento).

#### ***5.4.5 Evaluación del impacto***

Con el artículo 35 de la GDPR se incluye un análisis de riesgo que toda entidad sanitaria debe llevar a cabo con el objetivo de permitir a los responsables del tratamiento poder tomar las medidas oportunas para reducir los riesgos: La Evaluación del Impacto en la Protección de Datos (EIPD) [70].

Se trata de una metodología que permite evaluar anticipadamente las etapas en las que los datos son susceptibles a encontrarse bajo amenaza. Para el caso que nos ocupa se tendría que realizar de forma cíclica, pues en el campo sanitario se toman decisiones automatizadas y se tratan datos personales a gran escala, lo que se traduce a un riesgo para los derechos de los afectados.

Es por ello que esta metodología es constante y se constituye por la identificación de amenazas y la evaluación y tratamiento de riesgos (en caso de haberlos). Las medidas y protocolos de seguridad que se lleven a cabo se han de plasmar en un Documento de Seguridad a disposición de la AEPD para su consulta.

#### ***5.4.6 Comunicación de los datos personales***

La comunicación de datos entre entidades sanitarias no está destinada al interés legítimo de éstas para intereses laborales o comerciales. Nuestro caso se corresponde con la única excepción contemplada en la regulación, es decir, la comunicación de los datos entre una entidad clínica (mDurance) y una institución pública (Universidad Complutense de Madrid) cuyo proyecto de investigación en común tiene por objeto proporcionar facilidades a los expertos médicos en lo referente a la prevención, diagnóstico y asistencia sanitaria.

## Capítulo 6. Conclusiones

Por último, a modo de valoración global se destacan los aspectos fundamentales conseguidos con la realización de este documento. A modo de recapitulación, el autor de este Trabajo Final de Máster:

1. Tiene como origen de motivación los puntos comentados en el primer capítulo.
2. Se ha sumergido en un proceso de investigación y documentación sobre el marco teórico justificado en el capítulo 2.
3. Ha caracterizado mediante el tercer capítulo una metodología sobre la que realizar casos aplicables al objetivo general genérico propuesto en el proyecto.
4. Ha sido partícipe de un experimento perteneciente al ámbito de la Medicina del Deporte, el cual ha sido estudiado en el cuarto capítulo, y el que posee como forma de particularización la metodología propuesta.
5. Se ha documentado profundamente en la materia legislativa que implica manejar conjuntos de datos especialmente protegidos como son los que abarca el estudio práctico.

Todo esto ha permitido realizar un análisis completo sobre un área de estudio importante situada dentro del rango de potenciales que posee el mundo del Internet de las Cosas. Aquí, los conjuntos de datos procedentes de numerosas fuentes y dispositivos suponen un alto nivel de heterogeneidad en las integraciones e implican en muchas ocasiones una interpretación por parte del usuario final a modo de cajas negras. Es precisamente por esto que los sistemas Case-Based Reasoning juegan un papel importante en la generación de explicaciones, las cuales se establecen en una base de conocimiento que se retroalimenta de manera continua mediante la experiencia de casos anteriores.

En el caso que ocupa el proyecto, se ha comprobado que la metodología CBR, como conjunto de técnicas inteligentes, se adecúa muy correctamente a cualquier entorno IoT donde existan conjuntos de datos masivos que puedan ser incorporados a plantillas de informes recuperables y reutilizables. La metodología utilizada para ello es el propio ciclo CBR general mediante el cual cada una de las plantillas de informes tiene asociados unos elementos explicativos en forma de texto y gráficos. Ya se sabe que este Trabajo Final de Máster se engloba dentro de un proyecto de investigación con líneas de trabajo complementarias que se dedican a la implementación de dicho ciclo CBR general; y que el presente proyecto se dedica a plantear y diseñar explícitamente el sub-ciclo CBR que se encarga de la reutilización y adaptación de elementos gráficos.

Es por ello que, para la incorporación de elementos visuales en los informes, el sub-sistema CBR se ha de complementar con la ciencia DataViz, que supone un punto de inflexión en la capacidad de razonamiento humana ya que posibilita la generación de explicaciones visuales como forma de inteligencia. Es por ello que en el transcurso del proyecto se han estudiado las técnicas más importantes desde el punto de vista de la percepción humana y su relación con todos los atributos propios de los elementos gráficos.

En el proyecto se ha explicado un sistema genérico e independiente del dominio que sigue una metodología CBR aplicable a cualquier caso de uso donde prime la recuperación de plantillas de informes. Aquí se ha intensificado en el diseño de la fase del ciclo enfocada a la reutilización y adaptación de las visualizaciones, lo que supone

un sub-sistema CBR de recuperación de los gráficos candidatos al hueco de informe a completar, la correcta elección mediante adaptación, la revisión del gráfico elegido y la retroalimentación mediante la interacción que el usuario realiza con el gráfico.

Así pues, en lo que respecta a la materia práctica del proyecto, se ha llevado a cabo un experimento que supone la prueba de concepto de esta metodología. Este caso de estudio permite particularizar el sub-ciclo CBR de reutilización y adaptación de elementos DataViz en el dominio de la Medicina del Deporte, concretamente en el entorno de análisis de pruebas EMG, donde se permite al experto médico la evaluación del rendimiento de los músculos cervicales.

La forma en la que esta área de aplicación viene causando especial interés durante los últimos años en el mundo IoT se justifica con la constante incorporación de nodos y dispositivos encargados de la recogida de datos en este entorno, así como las múltiples soluciones de integración que estos wearables llevan de la mano. El desconocimiento por parte de los expertos hace que sientan real necesidad por disponer de informes donde se puedan interpretar estos datos de forma automática y con la mayor precisión posible, facilitando así la toma de decisiones en sus tareas.

Es por ello que mediante el experimento desarrollado en este Trabajo Final de Máster se ha conseguido diseñar e implementar un prototipo de herramienta de visualización de datos capaz de interpretar datos pertenecientes a este entorno específico de la Medicina del Deporte en base a unas determinadas reglas que atienden a cuatro modelos de informes considerados de utilidad en el estudio.

Para la utilización de los elementos gráficos correspondientes a cada tipo de informe se ha establecido una caracterización propia de cada modelo de informe que se sigue en todo momento para proporcionar una serie de elementos gráficos en función de la estructura de los datos propia de cada tipo de informe. Es importante decir que aquí existe una cierta acotación con respecto a los objetivos originales del proyecto, pues la recuperación del gráfico por lenguaje de petición de un experto médico se establece de forma limitada. Esto es, a pesar de que sí se ha implementado todo el conjunto de etiquetas que caracterizan tanto los datos como la base de conocimiento de los gráficos de forma totalitaria en los cuatro modelos de informes; la parte correspondiente al sistema de peticiones entre el médico y el sistema de reutilización de gráficos no ha quedado del todo completa, ya que únicamente se ha desarrollado una parte pormenorizada dentro del modelo de informe de inspección de un paciente individual.

Además, el prototipo de herramienta desarrollada ciertamente permite adelantarse en la detección de determinadas anomalías en función de las dependencias entre los datos estadísticos que se han etiquetado, pero únicamente desde el punto de vista de la analítico. Esto supone, por tanto, un tipo de acotación con respecto al tracking de esta interacción del experto médico con el gráfico para que se posea capacidad y autonomía en el proceso de aprendizaje.

Por tanto, se puede decir firmemente que sí se ha conseguido desarrollar un prototipo de herramienta que consigue facilitar la comprensión y visualización de los conjuntos de datos a los expertos médicos; así como favorecer la detección de determinadas anomalías desde la perspectiva analítica de los datos etiquetados en el sistema. No obstante, queda pendiente la ampliación del modelo de lenguaje común entre las peticiones de un experto médico y el sistema que recupera los gráficos, así como un tipo de soporte embebido que permita evaluar la interacción del experto con el elemento visual reutilizado y adaptado.

Por último, también se ha corroborado que la Medicina en general constituye un campo de gran relevancia en el ámbito IoT y en el tratamiento de datos masivo, ofreciendo una heterogeneidad especial e implicando un nivel de protección muy elevado. Es por este motivo por el cual se ha estudiado el Reglamento recientemente aplicable, que aumenta las obligaciones y responsabilidades para todos los procesos de tratamiento de datos, lo que supone una mejora en la restricción de los procedimientos propios de una entidad y en la comunicación de datos con otras entidades.

## **6.1 Limitaciones**

Este ambicioso proyecto supone una línea de investigación sobre un contexto en el que algunos procedimientos no poseen unas pautas firmemente preestablecidas. Por un lado, desde el punto de vista de la recogida y procesamiento de datos, tenemos que hoy día constantemente se están incorporando al mercado dispositivos IoT en este dominio específico que se encargan de recolectar datos y enviarlos a distintas plataformas sin seguir un patrón fijo. Esto es, la ingesta de los conjuntos cada vez se encuentra con una estructura distinta, llegando a tener los datos semi-estructurados o simplemente cadenas numéricas que llegan sin estructurar.

Por otro lado, también existe una limitación en el propio proceso de análisis de los datos EMG. En realidad, la detección de anomalías para el análisis EMG supone una fuente de discrepancias en determinados aspectos según la fuente de información que se consulte. Para la realización de este proyecto se han contrastado múltiples documentos de investigación y se han explicado los procesos que se consideran hoy en día más maduros en este aspecto.

Con respecto a la ejecución de tareas del proyecto en materia práctica, el principal inconveniente ha surgido en el abandono del corresponsable de la gestión del experimento tras la realización del mismo, por lo que a la hora de procesar, tratar, y estructurar los datos ha resultado algo más costoso y ha provocado que se emplee tiempo en unas tareas que podrían haber estado enfocadas en la optimización de la adaptación de los gráficos en el sub-ciclo CBR.

Un ejemplo de este tiempo empleado ha sido, sin ir más lejos, en la fase de procesado de los datos EMG, la cual ha sido un tanto tediosa sobre todo a la hora de determinar los tamaños de las ventanas para la correcta caracterización de los datos. Después, en la fase de estructuración y mapeado de las bases de datos, también ha surgido la dificultad de mapeado de los conjuntos de datos para establecer un patrón fijo, y es por ello que se ha estructurado primeramente en informes de población y paciente particular, para a posteriori establecer las comparaciones paciente-paciente y paciente-población.

## **6.2 Próximos pasos**

En primer lugar, como líneas de trabajo futuro enfocadas a la mejora de los procesos que engloba este proyecto, destacan dos aspectos importantes. Por un lado, tal y como se ha dicho, la optimización de la fase de recuperación y adaptación del gráfico en base a un lenguaje común entre el experto médico y el sistema para poder elegir correctamente el tipo de gráfico a representar, lo cual supone realizar una ampliación en los campos de relaciones entre las necesidades del experto médico acordes a cada uno de los cuatro modelos de informes.

Por otro lado, también se ha de optimizar el enfoque de la capacidad de aprendizaje de la herramienta, de modo que pueda disponer de aportaciones predictivas por parte del sistema para así ser capaz de aumentar su grado de personalización.

Ya se ha dicho que el entorno de trabajo utilizado para este proyecto es un banco de pruebas de Python. Una de las líneas de trabajo a seguir es la incorporación de este banco de pruebas a un entorno Cloud donde se pueda recoger datos de los sensores directamente, se puedan procesar, caracterizar y establecer un lenguaje común que procese las peticiones del experto médico en un entorno Python, para así enviarlas a Tableau, recuperar el dato, y devolverlo al entorno de la base de conocimiento.

## Conclusions

Finally, by way of an overall assessment, the fundamental aspects achieved with the production of this document are highlighted. As a summary, the author of this Final Master's Project:

1. He has as its source of motivation the points commented on in the first chapter.
2. He has immersed himself in a process of research and documentation on the theoretical framework justified in chapter 2.
3. By means of the third chapter, he has characterized a methodology on which to carry out cases applicable to the general objective proposed in the project.
4. He has participated in an experiment belonging to the field of Sports Medicine, which has been studied in the fourth chapter, and which has as a form of particularization the proposed methodology.
5. He has been deeply documented in the legislative matter that involves handling specially protected data sets such as those covered by the practical study.

All this has made it possible to carry out a complete analysis of an important area of study situated within the range of potentials possessed by the world of the Internet of Things. Here, data sets from numerous sources and devices imply a high level of heterogeneity in integrations and often imply an interpretation by the final user in the form of black-boxes. It is precisely for this reason that Case-Based Reasoning systems play an important role in the generation of explanations, which are established in a knowledge base that is continuously fed back through the experience of previous cases.

In the case occupied by the project, it has been verified that the CBR methodology, as a set of intelligent techniques, is very correctly adapted to any IoT environment where there are massive data sets that can be incorporated into recoverable and reusable report templates. The methodology used for this is the general CBR cycle itself, by means of which each of the report templates has associated explanatory elements in the form of text and graphics. It is already known that this Master's Final Project is part of a research main project with complementary lines of work that are dedicated to the implementation of this general CBR cycle; and that the present project is dedicated to explicitly raise and design the subcycle CBR that is responsible for the reuse and adaptation of graphic elements.

That is why, for the incorporation of visual elements in the reports, the CBR sub-system has to be complemented with the DataViz science, which represents a point of inflection in the human reasoning capacity since it makes possible the generation of visual explanations as a form of intelligence. During the project, the most important techniques have been studied from the point of view of human perception and their relationship with all the attributes of the graphic elements.

The project has explained a generic and domain independent system that follows a CBR methodology applicable to any case of use where the recovery of report templates prevails. The design of the part of the cycle focused on the reuse and adaptation of the visualizations has been intensified, which implies all the next: first, a CBR sub-system which recovers the graphics candidates to complete the report gap inside the template, secondly, the correct choice through adaptation, in third place, the revision of the chosen graphic, and finally, the feedback through the interaction that the user carries out with the graphic.

Thus, as far as the practical matter of the project is concerned, an experiment has been carried out which represents the proof of concept of this methodology. This case study allows for the particularization of the CBR subcycle for the reuse and adaptation of DataViz elements in the field of Sports Medicine, specifically in the EMG test analysis environment, where the medical expert is allowed to evaluate the performance of the cervical muscles.

The way in which this area of application has caused special interest in recent years in the IoT world is justified by the constant incorporation of nodes and devices responsible for data collection in this environment, as well as the multiple integration solutions that these wearables convey. The lack of knowledge on the part of the experts makes them feel a real need for reports where these data can be interpreted automatically and as accurately as possible, thus facilitating decision making in their tasks.

It is for this reason that by means of the experiment developed in this Final Master's Project it has been possible to design and implement a prototype of a data visualization tool capable of interpreting data belonging to this specific environment of Sports Medicine on the basis of certain rules that attend to four models of reports considered useful in the study.

In order to use the graphic elements corresponding to each type of report, a specific characterization of each report model has been established and followed in order to provide a series of graphic elements depending on the structure of the data specific to each type of report. It is important to say that here there is a certain restriction regarding the original objectives of the project, since the recovery of the graphic by request language of a medical expert is established in a limited way. That is, despite the fact that all the set of labels which characterize both data and knowledge base (created from graphs) have been implemented in a totalitarian way in the four report models; the respective part of requests language between the doctor and the system of graphs' reuse has not been fully completed, as only a detailed part has been developed within the inspection report model of an individual patient.

In addition, the tool prototype developed certainly makes it possible to advance in the detection of certain anomalies depending on the dependencies between the statistical data that have been labelled, but only from the analytical point of view. This implies, therefore, a type of restriction in regards to the tracking of this interaction between the medical expert and the graphic, in order to have capacity and autonomy at the learning process.

Therefore, it can be firmly said that it has been possible to develop a tool prototype that makes it easier for medical experts to understand and visualize the data sets, as well as favoring the detection of certain anomalies from the analytical perspective of the data labelled at the system. However, two kind of work streams remain to be extended. Firstly, the common language model between the requests of a medical expert and the system that retrieves the graphics, and secondly, a type of embedded support that allows evaluating the interaction of the expert with the reused and adapted visual.

Finally, it has also been checked that Medicine in general represents a field of great relevance in IoT and also in the treatment of massive data, offering a special heterogeneity and implying a high level of protection. It is for this reason that the recently applicable Regulation has been studied, which increases the obligations and responsibilities for all data processing processes, and of course means an improvement in the restriction of the procedures of one entity and in the communication of data with other entities.

## **Limitations**

This ambitious project involves a research line on a context in which some procedures do not have firmly pre-established guidelines. On the one hand, from the point of view of data collection and processing, nowadays we are constantly incorporating into the IoT market multiple devices in this specific domain that are responsible for collecting data and sending them to different platforms without following a fixed pattern. This is, the ingestion of the sets each time is with a different structure, coming to have the data semi-structured or simply unstructured strings.

On the other hand, there is also a limitation in the EMG data analysis process itself. Indeed, the detection of anomalies for EMG analysis is a source of discrepancies in certain aspects depending on the source of information consulted. In order to appropriately implement this project, multiple research documents have been contrasted and the processes that are currently considered more mature in this aspect have been explained.

With regard to the execution of practical project tasks, the main drawback has arisen in the abandonment of the co-manager of the experiment after its completion, so that when processing, treating, and structuring the data has been much more difficult and has led to spending time on tasks that could have been focused on optimizing the adaptation of graphics in the subcycle CBR.

An example of this time spent has been in the phase of processing EMG data, which has been somewhat tedious especially when determining the sizes of the windows for the correct processing and characterization of the data. Later, in the phase of structuring and mapping databases, has also arisen the difficulty of mapping the data sets to establish a fixed pattern, and that is why it has been structured firstly in reports of population and particular patient, to later establish the patient-patient and patient-population comparisons.

## **Next steps**

Firstly, as lines of future work focused on improving the processes covered by this project, two important aspects stand out. On the one hand, as it has been said, the optimization of the recovery phase and adaptation of the graph based on a common language between the medical expert and the system in order to be able to choose correctly the type of graph to be represented, which involves an expansion in the relations between the needs of the medical expert according to each of the four report models.

On the other hand, it is also necessary to optimize the approach of the learning capacity of the tool, so that it can provide predictive contributions from the system and be able to increase its degree of personalization.

It has already been said that the working environment used for this project is a Python test bench. One of the lines of work to be followed is the incorporation of this test bench into a Cloud environment where sensor data can be collected directly, processed, characterized and established a common language that processes the requests of the medical expert in a Python environment, in order to send them straightaway to Tableau, retrieve the data, and return it to the knowledge base environment.

## Bibliografía

- [1] «Inteligencia artificial: orígenes y evolución de las máquinas inteligentes», *TheDailyProsper*. [En línea]. Disponible en: <https://thedailyprosper.com/es/a/inteligencia-artificial-origenes-y-evolucion-de-las-maquinas-inteligentes>. [Accedido: 28-feb-2018].
- [2] O’Leary y Daniel E., «Artificial intelligence and big data», *IEEE Intell. Syst.*, vol. 28, n.º 2, pp. 96–99, 2013.
- [3] Belén Díaz-Agudo, Juan A. Recio-García, y Guillermo Jimenez-Díaz, «Data explanation with CBR», *Dep. Softw. Eng. Artif. Intell.*, vol. Universidad Complutense de Madrid, Spain.
- [4] «What is data quality?», *SearchDataManagement*. [En línea]. Disponible en: <https://searchdatamanagement.techtarget.com/definition/data-quality>. [Accedido: 03-jun-2018].
- [5] Edward Segel y Jeffrey Heer, «Narrative Visualization: Telling Stories with Data», *IEEE Trans. Vis. Comput. Graph.*, vol. VOL. 16, n.º NO. 6, dic. 2010.
- [6] «La carrera por la industria de la inteligencia artificial», *ELESPECTADOR*. [En línea]. Disponible en: <https://www.elespectador.com/tecnologia/la-carrera-por-la-industria-de-la-inteligencia-artificial-articulo-755627>. [Accedido: 03-mar-2018].
- [7] R. García-Retamero y E. T. Cokely, «Communicating health risks with visual aids», *Curr. Dir. Psychol. Sci.*, vol. 22, n.º 5, pp. 392–399, 2013.
- [8] Ricardo Barranco Frago, «¿Qué es Big Data?», *IBM*. [En línea]. Disponible en: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>. [Accedido: 21-mar-2018].
- [9] «Big Data y GIS», *Geodan*, 03-feb-2015. [En línea]. Disponible en: <https://www.geodan.es/big-data-y-gis/>. [Accedido: 11-abr-2018].
- [10] «IoT, Big Data and AI – the New ‘Superpowers’ In the Digital Universe», *Business 2 Community*. [En línea]. Disponible en: <https://www.business2community.com/big-data/iot-big-data-ai-new-superpowers-digital-universe-01926411>. [Accedido: 07-jun-2018].
- [11] Departamento de Ingeniería del Software e Inteligencia Artificial, «Introducción a SBC - Inteligencia Artificial». Universidad Complutense Madrid, Curso /2018-2017.
- [12] Ashok K Goel y Belén Díaz-Agudo, «What’s hot in CBR». Association for the Advancement of Artificial Intelligence, 2017.
- [13] Agnar Aamodt y Enric Plaza, «Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches», *AI Commun. IOS Press*, vol. Vol. 7, n.º 1, pp. 39-59.
- [14] «Analysts: Data visualization tools key to “big data” analytics success», *SearchBusinessAnalytics*. [En línea]. Disponible en: <https://searchbusinessanalytics.techtarget.com/news/2240111798/Analysts-Data-visualization-tools-key-to-big-data-analytics-success>. [Accedido: 18-mar-2018].
- [15] «Preattentive Visual Properties and How to Use Them in Information Visualization», *The Interaction Design Foundation*. [En línea]. Disponible en: <https://www.interaction-design.org/literature/article/preattentive-visual-properties-and->

- how-to-use-them-in-information-visualization. [Accedido: 04-may-2018].
- [16] Stephen Few, *Data Visualization for Human Perception*, 2nd ed. .
- [17] «Design Principles: Visual Perception And The Principles Of Gestalt», *Smashing Magazine*. [En línea]. Disponible en: <https://www.smashingmagazine.com/2014/03/design-principles-visual-perception-and-the-principles-of-gestalt/>. [Accedido: 13-abr-2018].
- [18] Stephan Few, *Information Dashboard Design: The Effective Visual Communication of Data*, O'REILLY. 2006.
- [19] «Choosing colors for your data visualization», *Cambridge Intelligence*, 01-nov-2016. [En línea]. Disponible en: <https://cambridge-intelligence.com/choosing-colors-for-your-data-visualization/>. [Accedido: 28-may-2018].
- [20] «An introduction to the HSL color system», *Nix Sensor Ltd*, 14-feb-2018. .
- [21] «Color wheel | Color schemes - Adobe Color CC», *Adobe Color*. [En línea]. Disponible en: <https://color.adobe.com/create/color-wheel/?base=2&rule=Analogous&selected=3&name=My%20Color%20Theme&mode=rgb&rgbvalues=1,0.3270833333331894,0.050000000000000044,0.91,0.17157291666660116,0.04550000000000004,1,0,0,0.91,0.04550000000000004,0.4777500000001965,1,0.050000000000000044,0.999999999999784&swatchOrder=0,1,2,3,4>. [Accedido: 30-may-2018].
- [22] Edward R. Tufte, *The Visual Display of Quantitative Information*, vol. Graphic Press. Cheshire, Connecticut.
- [23] «¿Qué es una infografía? - Pixel Creativo». [En línea]. Disponible en: <https://pixel-creativo.blogspot.com/2012/01/que-es-una-infografia.html>. [Accedido: 20-jun-2018].
- [24] J. Steele y N. P. N. Iliinsky, Eds., *Beautiful visualization: looking at data through the eyes of experts*, 1st ed. Sebastopol, CA: O'Reilly, 2010.
- [25] «Five Tips for Nonprofits Data Visualization | Beth's Blog», *Bethkanter*. [En línea]. Disponible en: <http://www.bethkanter.org/data-viz-2/>. [Accedido: 15-may-2018].
- [26] «5 Amazing Ways to Impact Your Audience With Data Storytelling», *SnapApp*, 21-dic-2017. [En línea]. Disponible en: <https://www.snapapp.com/blog/5-amazing-ways-impact-your-audience-data-storytelling/>. [Accedido: 30-mar-2018].
- [27] «Great Presentations Incorporate Story», *Duarte*, 25-abr-2014. .
- [28] J. F. Martinez, «Datos personales sanitarios: qué son y cómo tratarlos correctamente», *clickDatos Protección de Datos*, 23-feb-2018. [En línea]. Disponible en: <https://clickdatos.es/tratar-datos-personales-sanitarios/>. [Accedido: 16-may-2018].
- [29] «Electromyography (EMG): Purpose, Procedure, and Results», *Healthline*, 20-mar-2018. [En línea]. Disponible en: <https://www.healthline.com/health/electromyography>. [Accedido: 23-jun-2018].
- [30] «Unidad de Medicina del Deporte - Rehabilitación», *Sanitas*. [En línea]. Disponible en: [//www.sanitas.es/sanitas/seguros/es/particulares/servicios\\_salud/rehabilitacion/unidad-medicina-deporte/index.html](http://www.sanitas.es/sanitas/seguros/es/particulares/servicios_salud/rehabilitacion/unidad-medicina-deporte/index.html). [Accedido: 24-abr-2018].
- [31] «Ministerio de Sanidad, Servicios Sociales e Igualdad - Ciudadanos - Sistema

- Nacional de Salud - Hospitales - Catálogo Nacional de Hospitales», *MSSSI*. [En línea]. Disponible en: <https://www.msssi.gob.es/ciudadanos/prestaciones/centrosServiciosSNS/hospitales/introduccion.htm>. [Accedido: 25-abr-2018].
- [32] Redacción, «Tratamiento de datos de big data en en el Internet de las Cosas», *PowerData*. [En línea]. Disponible en: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/tratamiento-de-datos-de-big-data-en-en-el-internet-de-las-cosas>. [Accedido: 24-may-2018].
- [33] «¿Qué es IoMT (internet de las cosas médicas) o IoT de salud?», *SearchDataCenter*. [En línea]. Disponible en: <https://searchdatacenter.techtarget.com/es/definicion/IoMT-internet-de-las-cosas-medicas-o-IoT-de-salud>. [Accedido: 08-abr-2018].
- [34] «¿Por qué se debe implementar la narrativa visual en las grandes empresas?», *Reporte Digital*, 07-jun-2016. .
- [35] S. F. Roth y J. Mattis, «Data characterization for intelligent graphics presentation», en *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1990, pp. 193–200.
- [36] S. F. Roth y J. Mattis, «Automating the Presentation of Information», 1991.
- [37] M. X. Zhou y S. K. Feiner, «Data characterization for automatically visualizing heterogeneous information», en *Information Visualization'96, Proceedings IEEE Symposium on*, 1996, pp. 13–20.
- [38] «Estadística para todos», *Estadística para todos*. [En línea]. Disponible en: <http://www.estadisticaparatodos.es/historia/historia.html>. [Accedido: 15-jul-2018].
- [39] Plácido Guardiola, «Parámetros y Tipos de Estadísticos». Universidad de Murcia.
- [40] J. A. Fernández-Niño, R. I. Hernández-Montes, y L. A. Rodríguez-Villamizar, «Reporte estadístico en los análisis de regresión en Biomédica: una revisión y evaluación crítica», *Biomédica*, vol. 38, n.º 2, pp. 173-179, jun. 2018.
- [41] J. Zubcoff, «Regresión lineal», *Introd. Inferencia Estad.*, 2008.
- [42] «Algunos métodos simples para la evaluación de recursos pesqueros tropicales», *Fao*. [En línea]. Disponible en: <http://www.fao.org/docrep/003/x6845s/x6845s02.htm>. [Accedido: 20-jul-2018].
- [43] I. M. Peláez, *Modelos de regresión: lineal simple y regresión logística*. SEDEN, 2006.
- [44] W. Spy, «Thick Data - Word Spy». [En línea]. Disponible en: <https://wordspy.com/index.php?word=thick-data>. [Accedido: 07-jul-2018].
- [45] «The Power of Thick Data | BIGfish PR», *BIGfish PR | Tech PR Agency, Public Relations Firm*. .
- [46] Logicalis, «Visualización de datos inteligente», *Logicalis*. [En línea]. Disponible en: <https://blog.es.logicalis.com/analytics/visualizacion-de-datos-inteligente>. [Accedido: 21-jul-2018].
- [47] «The Data Visualisation Catalogue», *DataViz Cataogue*. [En línea]. Disponible en: <https://datavizcatalogue.com/>. [Accedido: 27-jun-2018].

- [48] «mDurance», *mDurance*. [En línea]. Disponible en: <http://www.mdurance.eu/>. [Accedido: 12-feb-2018].
- [49] A. D. Nimbarte, M. Zreiqat, y X. Ning, «Impact of shoulder position and fatigue on the flexion–relaxation response in cervical spine», *Clin. Biomech.*, vol. 29, n.º 3, pp. 277-282, mar. 2014.
- [50] R. Mousavi-Khatir, S. Talebian, N. Maroufi, y G. R. Olyaei, «Effect of static neck flexion in cervical flexion-relaxation phenomenon in healthy males and females», *J. Bodyw. Mov. Ther.*, vol. 20, n.º 2, pp. 235-242, abr. 2016.
- [51] C. F. Pinheiro, M. F. dos Santos, y T. C. Chaves, «Flexion–relaxation ratio in computer workers with and without chronic neck pain», *J. Electromyogr. Kinesiol.*, vol. 26, pp. 8-17, feb. 2016.
- [52] L. Mi-ra, Y. Won-gyu, A. Duk-hyun, K. Mi-hyun, y O. Jae-seop, «The Effect of Backpack Loads on FRR (Flexion-relaxation Ratio) in the Cervical Spine», mar. 2011.
- [53] Shimmer, «Wearable Sensor Technology | Wireless IMU | ECG | EMG | GSR», *Shimmer Sensing*. [En línea]. Disponible en: <http://www.shimmersensing.com/>. [Accedido: 07-ago-2018].
- [54] «Project Jupyter», *Jupyter*. [En línea]. Disponible en: <http://www.jupyter.org>. [Accedido: 01-mar-2018].
- [55] «Tableau Software», *Tableau Software*. [En línea]. Disponible en: <https://www.tableau.com/>. [Accedido: 06-mar-2018].
- [56] J.-P. Pialasse, D. Lafond, V. Cantin, y M. Descarreaux, «Load and speed effects on the cervical flexion relaxation phenomenon», *BMC Musculoskelet. Disord.*, vol. 11, n.º 1, dic. 2010.
- [57] «¿Qué son los datos de carácter personal? - LOPD», *Cuida tus datos*. [En línea]. Disponible en: <http://www.cuidatusdatos.com/infodatospersonales.html>. [Accedido: 24-abr-2018].
- [58] AEPD, «Agencia Española de Protección de Datos», *Portal AGPD*. [En línea]. Disponible en: <https://www.agpd.es/portalwebAGPD/CanalDelCiudadano/derechos/index-ides-idphp.php>. [Accedido: 17-abr-2018].
- [59] «Título I. De los derechos y deberes fundamentales - Constitución Española», *Congreso*. [En línea]. Disponible en: <http://www.congreso.es/consti/constitucion/indice/titulos/articulos.jsp?ini=18&tipo=2>. [Accedido: 27-abr-2018].
- [60] Jefatura del Estado, «Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal», *Bol. Of. Estado BOE*, n.º COMENTARIO A LA LEY DE PROTECCIÓN DE DATOS, p. 41, 2010.
- [61] «¿Qué hay que hacer para cumplir con la LOPD?», *Cuida tus datos*. [En línea]. Disponible en: <http://cuidatusdatos.com/infocomocumplirlopd.html>. [Accedido: 11-may-2018].
- [62] «REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de

Datos)», *Unión Eur.*

[63] «Ley de Protección de Datos en sanidad: guía recomendada para médicos», *Protección datos LOPD*. [En línea]. Disponible en: <https://www.protecciondatos-lopd.com/empresas/guia-centros-sanitarios/>. [Accedido: 12-may-2018].

[64] «Artículo 9 del Reglamento General de Protección de Datos», *Unión Eur.*

[65] Herrero Abogados, «Los centros de salud y el nuevo Reglamento General de Protección de Datos», *Herrero Digital*. [En línea]. Disponible en: <http://www.herrerodigital.com/blog/los-centros-de-salud-y-el-nuevo-reglamento-general-de-proteccion-de-datos/>. [Accedido: 17-abr-2018].

[66] «Delegado de Protección de Datos DPO, funciones certificación APEP», *Ayuda Ley Protección Datos (LOPD)*, 27-jun-2016. [En línea]. Disponible en: <https://ayudaleyprotecciondatos.es/2016/06/27/delegado-proteccion-datos/>. [Accedido: 29-abr-2018].

[67] «GDPR: 7 claves para hospitales y clínicas», *Signaturit*. [En línea]. Disponible en: <https://blog.signaturit.com/es/gdpr-claves-para-hospitales-y-clinicas>. [Accedido: 12-abr-2018].

[68] AEPD y ISMS, «Código de buenas prácticas en protección de datos para proyectos BIG DATA», p. 40.

[69] IT Digital Media, «Cómo mitigar los ciberriesgos de los dispositivos de IoT médicos | Vulnerabilidades», *IT Digital Security*, 02-ene-2018. [En línea]. Disponible en: <http://www.itdigitalsecurity.es/vulnerabilidades/2018/01/como-mitigar-los-ciberriesgos-de-los-dispositivos-de-iot-medicos>. [Accedido: 22-may-2018].

[70] J. A. Prego de Oliver Fernández, «La transparencia como elemento de apoyo al consentimiento en materia de protección de datos», *UC3M*, 2017.

[71] Shimmer, «Shimmer ECG EMG User Guide», 2017. [En línea]. Disponible en: [www.shimmersensing.com](http://www.shimmersensing.com). [Accedido: 04-ago-2018].

## Anexos

### Anexo 1: Especificaciones del wearable Shimmer 3 EMG [71]

El wearable Shimmer3 EMG es capaz de medir y grabar la actividad eléctrica asociada a las contracciones de los músculos del paciente. Esto es una fuente de datos muy valiosa en el ámbito biomédico ya que permite a los expertos evaluar la conducción nerviosa de músculos categorizados como totalmente sanos y permitir contrastarla con tejidos musculares lesionados.

Este dispositivo tiene el aspecto físico que se muestra en la figura 84:



Figura 84. Wearable Shimmer3 EMG, encargado de recoger las señales musculares del experimento.

El Shimmer3 EMG realiza la función de sensado mediante cinco canales que se corresponden con cinco electrodos de superficie. Los canales uno y dos tienen dos tomas de contacto cada uno, mientras que la toma de tierra equivale a un único electrodo. Así pues, ambos canales poseen una protección de desfibrilación sobre el canal 1 y el de referencia, excluyendo las tomas del canal 2 para facilitar la demodulación.

Esto se puede observar en el diagrama de la figura 85, donde se muestran los elementos básicos del dispositivo:

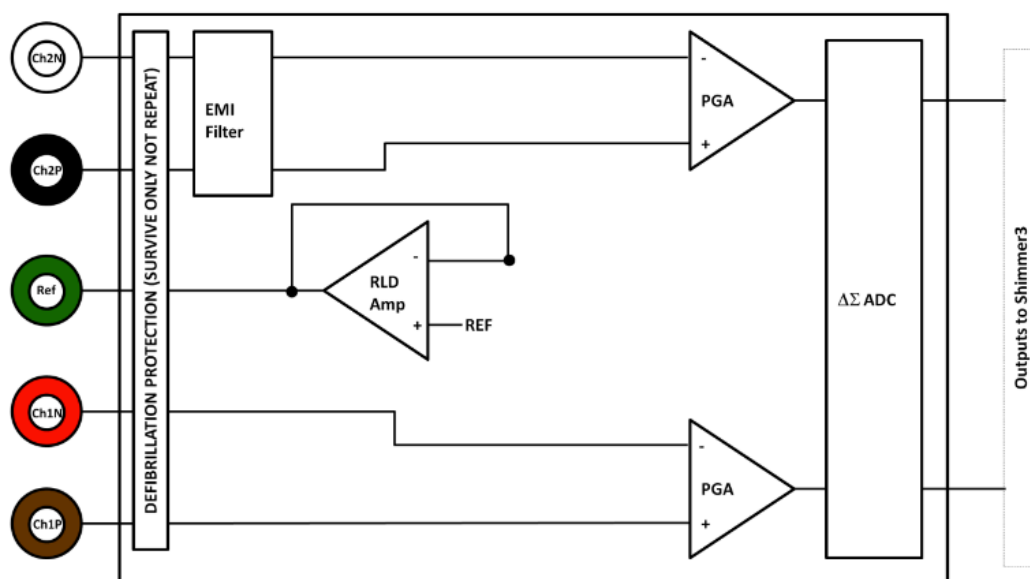


Figura 85. Diagrama de bloques simplificado del wearable Shimmer3.

Así pues, se observan tres elementos destacables:

- En las tomas del canal 2 se establece un *filtro EMI* que reduce la interferencia electromagnética con un filtro de aproximadamente 3MHz de ancho de banda.
- El *amplificador RLD* de la toma de referencia se encarga de contrarrestar las interferencias comunes como líneas de alimentación y otras fuentes.
- El amplificador PGA incrementa la amplitud de las señales de entrada con siete opciones de regeneración programables.
- La toma de salida se compone de un convertidor A/D mediante una representación de 24 bits por cada muestra de señal tomada. Estos valores se transfieren al procesador del Shimmer3 para ser almacenados en el dispositivo o transmitidos mediante el protocolo de comunicación.

Por último, en la tabla siguiente se muestra un resumen de las especificaciones técnicas más relevantes de este dispositivo en cuanto a procesamiento, capacidad de almacenamiento, protocolo de comunicación, alimentación, y otros componentes que resultan de interés.

Capacidad	Detalle
Procesamiento	Microcontrolador MSP 430 (24MHz, 16 bit)
Comunicación	Bluetooth –RN42, Radio: 802.15.4 – TI CC2420
Almacenamiento	Lector de tarjetas microSD 8GB integrado
Batería	450mAh recargable Li-ion
Acelerómetro de tres ejes integrados	Freescale MMA7361
Rango acelerómetro	+/- 1.5g – 6g
Ganancia	Configurable (1, 2, 3, 4, 6, 8, 12)

Tasa de datos	Software configurable (125, 250, 500, 1000, 2000, 4000, 8000 SPS)
Rango dinámico de entrada diferencial	Aproximadamente 800 mV (para ganancia = 6)
Ancho de banda	8.4 kHz
Tierra	Tipo Wilson
Protección de entrada	Filtrado ESD Y RF/EMI; Limitante de corriente; Entradas incluyen protección de desfibrilación
Conexiones	Entrada Canal1N, Entrada Canal1P, Entrada Canal2N, Entrada Canal2P, Referencia (Ref)
Peso	31 gramos
Dimensiones	65 x 32 x 12 mm
Memoria EEPROM	2048 bytes

**Tabla 13. Especificaciones técnicas del wearable Shimmer3.**

## Anexo 2: Paletas de color utilizadas

En las siguientes figuras se detallan las paletas de color utilizadas en el diseño de los gráficos, según funcionalidad.

Por un lado, para las etiquetas que describen características propias de cada paciente, como el sexo, la edad, o el propio identificador de paciente, se utilizan dos tonos de color complementarios entre sí:



Figura 86. Paleta de colores número 1.

Por otro lado, para las etiquetas semánticas enfocadas a las alteraciones o dolencias de los pacientes de la población se hace uso de colores análogos jugando con distintos niveles de contraste:

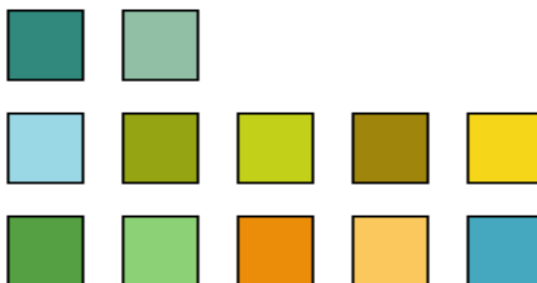


Figura 87. Paleta de colores número 2.

Por último, para las etiquetas sintácticas en la representación de la actividad neuromuscular se utilizan colores complementarios entre sí, dentro de los siguientes tonos:



Figura 88. Paleta de colores número 3.