



Natural Language and AI solutions for citizen services evolution

Salvador Estevan Martínez

Deputy Head of Unit of Communications

General Secretariat for Digital Administration

Doaa Samy

Advanced Computational Linguist

General Technical Office

Spanish Plan for Advancement of Language Technologies

- I. **Introduction.**
- II. AI. Machine Learning. NLP.
- III. What are Public Services doing?
- IV. 060 Service
- V. Pilots. Chatbot –virtual assistant.
- VI. Work in progress.

Introduction

✓ Some facts:

- Users and citizen are demanding an immediate attention, omni-channeling, 24x7.
- This type of attention is increasing. Cause: digital users.
- More than the 70% of 'Millennial' avoid human interaction, and prefer self-service solutions, as an American investigation states.
- Conversational interfaces use is growing: Siri, Google Assistant, Alexa, etc.
- Financial, IT, insurances... companies are developing chatbots for the customer service.
- The European Commission is about to invest 1.500 millions € in Artificial Intelligence for the period 2018-2020, under the scope of the Program Horizon 2020. For the period of 2021-2027: will be investing at least 7.000 millions €.

Introduction

✓ Some facts:

- Users and citizen are demanding an immediate attention, omni-channeling, 24x7.
- This type of attention is increasing. Cause: digital users.
- More than the 70% of 'Millennial' avoid human interaction, and prefer self-service solutions, as an American study states.
- Conversational interfaces use is growing: Siri, Google Assistant, Alexa, etc.
- Financial, IT, insurances... companies are developing chatbots for the customer service.
- The European Commission is about to invest 1.500 millions € in Artificial Intelligence for the period 2018-2020, under the scope of the Program Horizon 2020. For the period of 2021-2027: will be investing at least 7.000 millions €.

**Technology for a better use of resources, more efficient
Seeking value**

- I. Introduction.
- II. **AI. Machine Learning. NLP.**
- III. What are Public Services doing?
- IV. 060 Service
- V. Pilots. Chatbot –virtual assistant.
- VI. Work in progress.

IA. Machine Learning. NLP.



Artificial
Intelligence

✓ Technology that allows a computer to simulate human intelligence. It does this by taking in information from its surroundings, and deciding its response based on what it learns or senses. It acquires capabilities like reasoning, learning or self-correction.

- Conversational AI.



Machine Learning

✓ **Machine Learning** is an application of the **Artificial Intelligence** based on: granting machines access to data and let them learn for themselves.

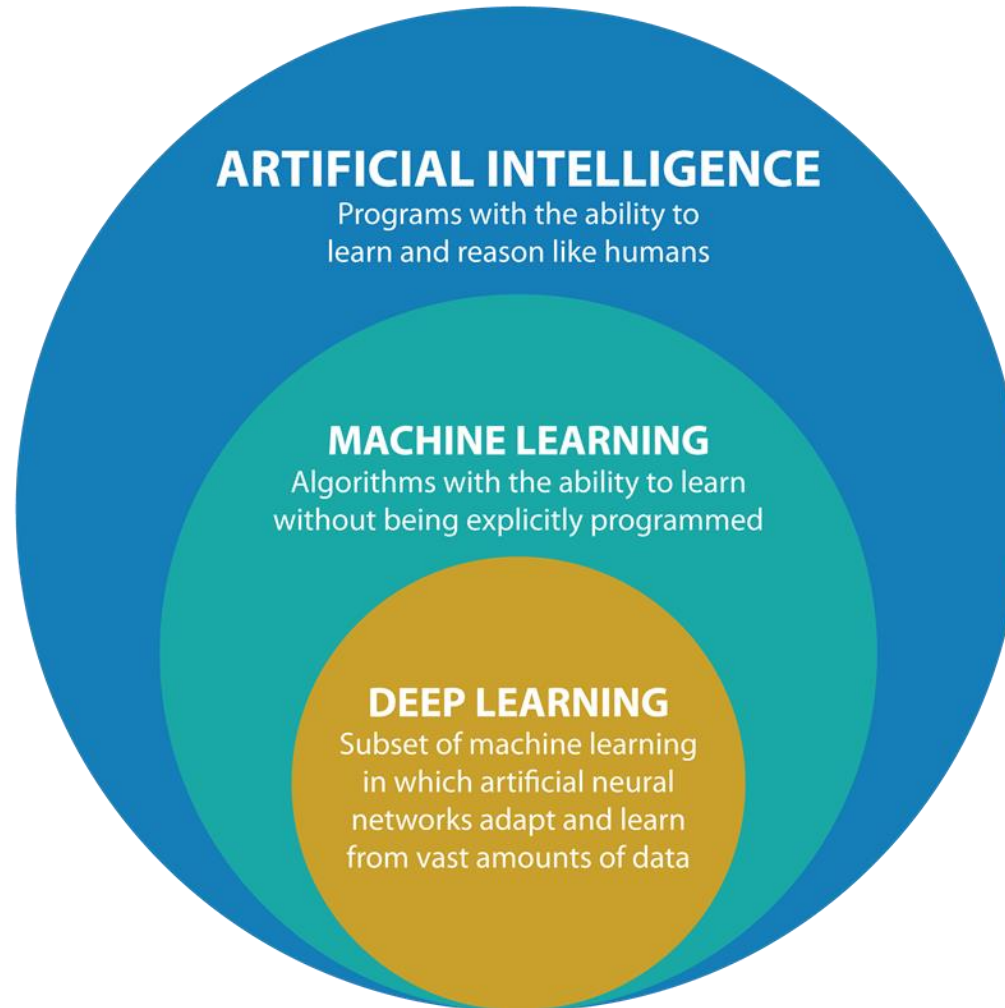


Human Language
Technologies

✓ **IA** application to **NLP**:

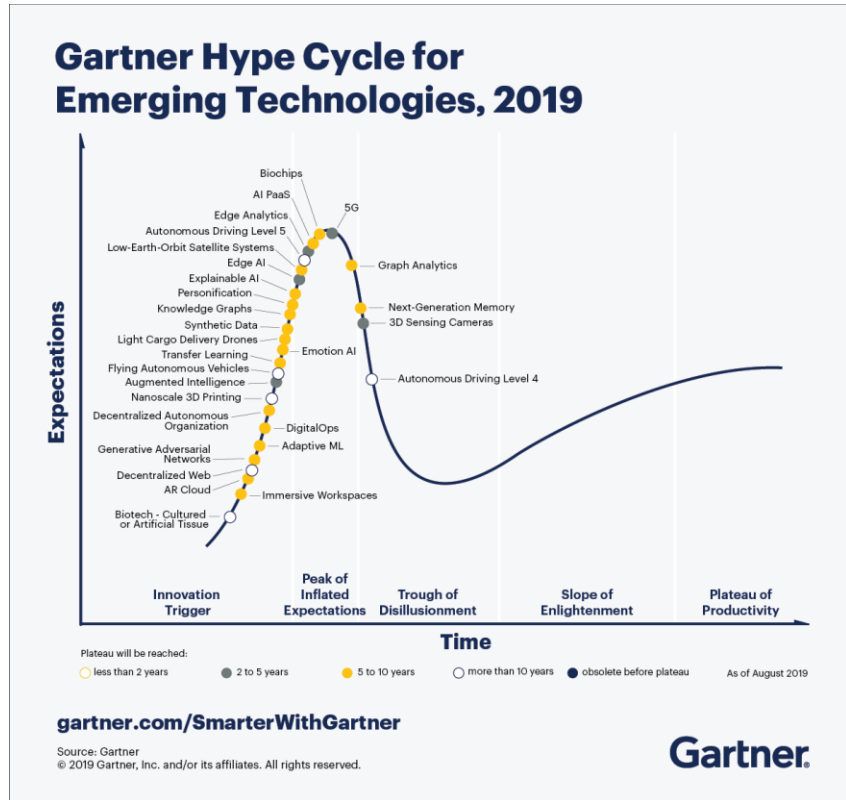
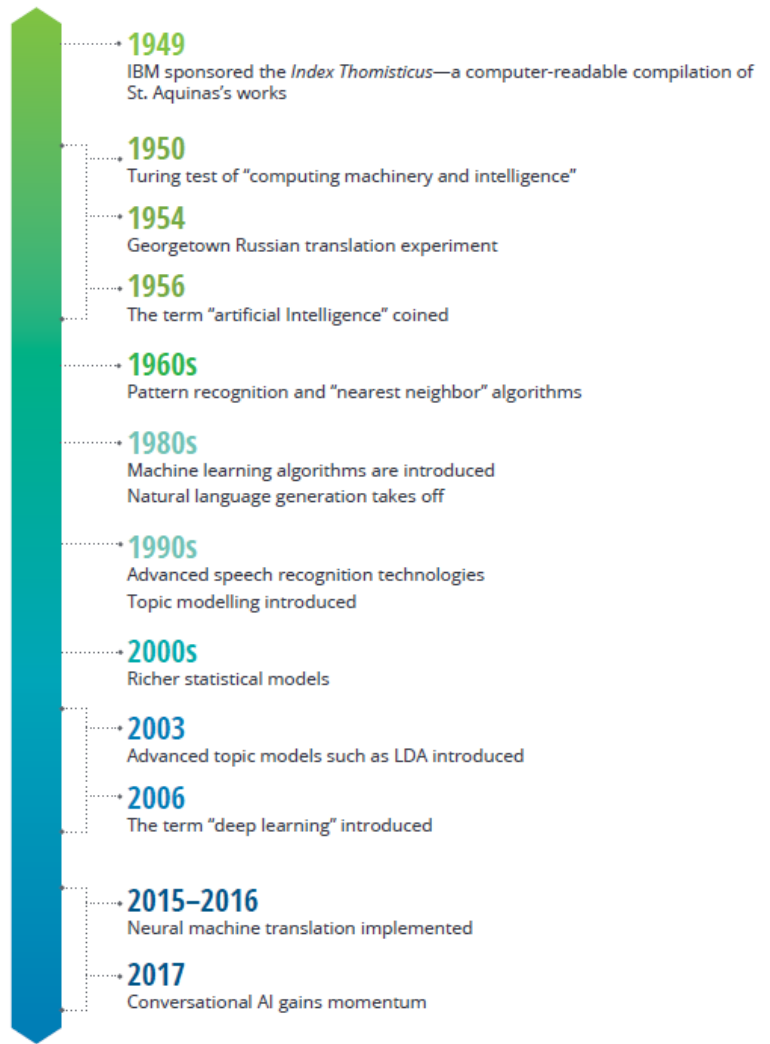
- Language detection. Relationship identification.
- Content categorization.
- Insights into customer sentiment.
- Summarize information.
- Translation.

IA. Machine Learning. NLP.



IA. Machine Learning. NLP.

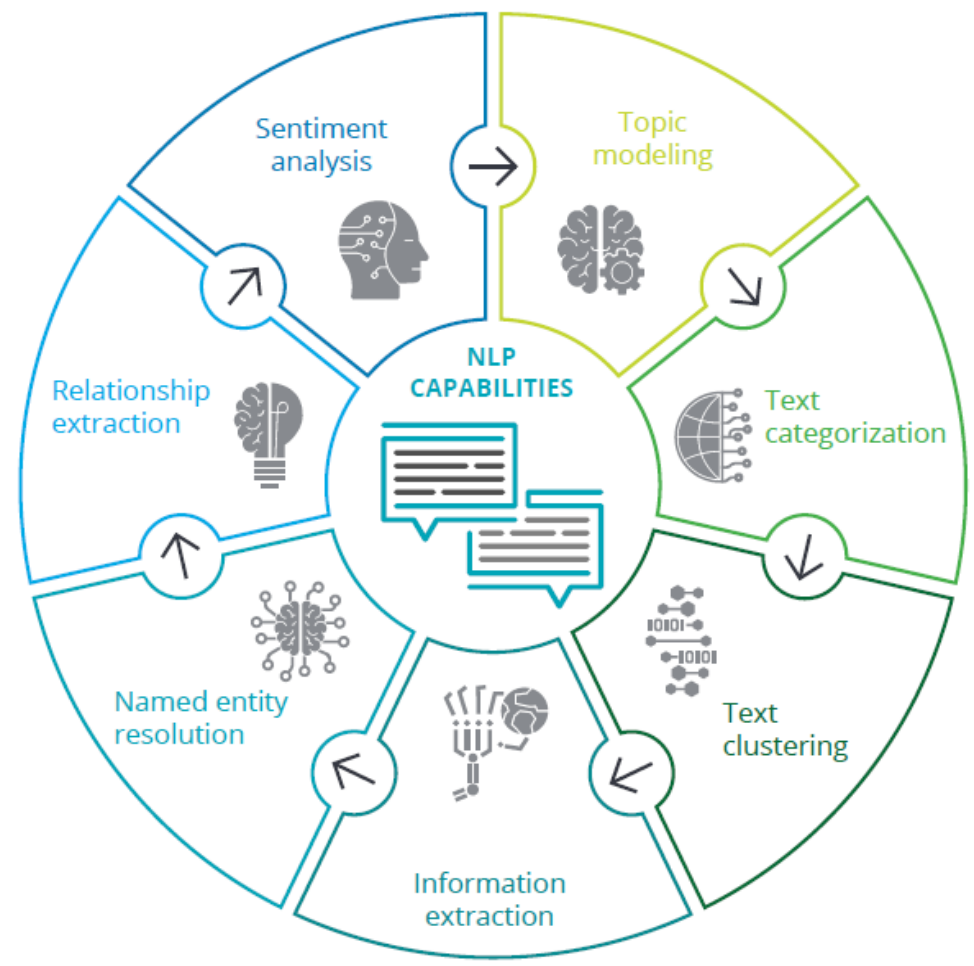
The evolution of NLP and underlying algorithms



Sources: Thomas N. Winter, "Roberto Busa, S.J., and the inventor of the machine-generated concordance," University of Nebraska-Lincoln, January 1999; Bhargav Shah, "The power of natural language processing: Today's boom in artificial intelligence," *Medium*, July 13, 2017; Chris Smith et al., *The history of artificial intelligence*, University of Washington, December 2006; Eric Eaton, "Introduction to machine learning," presentation, University of Pennsylvania; Kendall Fortney, "Pre-processing in natural language machine learning," *Towards Data Science*, November 28, 2017; Clark Boyd, "The past, present, and future of speech recognition technology," *Medium*, January 10, 2018; Regina Barzilay, "Natural language processing," lecture, MIT, September 2012; Build with Google Cloud, "A history of machine learning," accessed December 19, 2018; Thomas Hofmann, "Probabilistic latent semantic indexing," proceedings of the twenty-second Annual International SIGIR Conference on Research and Development in Information Retrieval, accessed December 19, 2018; Robert Dale, Barbara Di Eugenio, and Donia Scott, "Introduction to the special issue on natural language generation," *Computational Linguistics* 24, no. 3 (September 1998); *Medium*, "History and frontier of the neural machine translation," August 17, 2017; Ram Menon, "The rise of the conversational AI," *Forbes*, December 4, 2017.

IA. Machine Learning. NLP.

Key NLP capabilities



Source: Deloitte analysis.

So, what about the Public Administration?

- I. Introduction.
- II. IA. Machine Learning. NLP.
- III. **What are Public Services doing?**
- IV. 060 Service
- V. Pilots. Chatbot –virtual assistant.
- VI. Work in progress.

Guidelines

Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje



1. Why Language Technologies? -> Introduction
2. How is *Plan TL* structured? -> Framework of Plan TL
3. Where does *Semantic Interoperability of Metadata* lie within Plan TL?
4. What are the results of Plan? -> Use cases: Dialogue Systems, Machine Translation and Competitive Intelligence

1.1. What are Human Language Technologies (HLT)?

- **Human Language Technologies** are the set of technologies that enable to interface with machines by voice and language.
 - HLT enable the extraction of meaning from data, turning it into useful knowledge.
 - Tools and services to analyse both structured (text, documents) and unstructured data (human speech, social media content) are required in order to fully exploit the huge quantities of data available.
- The full range of topics covered are:
 - Machine Translation,
 - Multilingual content authoring and management,
 - Speech Technology and interactive services,
 - Content Analytics,
 - Language Resources
 - <https://ec.europa.eu/digital-single-market/en/programme-and-projects/language-technologies-project-information>

1.2. Why Human Language Technologies?

- HLT are essential for **multilingual** Europe: cross-lingual communication and cross-border information exchange is of fundamental importance.
- Crucial component for Artificial Intelligence and Disruptive Technologies.
- Enabler for **High Growth Innovative Industries** in the Digital Single Market.
- HLT can help in offering new **public services** for citizens and **enterprises** on strategic sectors (health, justice, etc.).
- High potential for **internationalization** of the **Spanish language** and strong cooperation with Latin America.

Guidelines

Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje



1. Why Language Technologies? -> Introduction
2. How is *Plan TL* structured? -> Framework of Plan TL
3. Where does *Semantic* Interoperability of *Metadata* lie within Plan TL?
4. What are the results of Plan? -> Use cases: Dialogue Systems, Machine Translation and Competitive Intelligence

2.1. Framework of Plan TL: Cross-cutting Technical Lines

1. Language Resources

2. Techniques and Algorithms (Machine Learning-Deep Learning)

3. Natural Language Processing (NLP)

4. Machine Translation

5. Speech Technologies and Dialogue Systems

2.2. Framework of Plan TL: Action Lines

1. Linguistic infrastructures development

2. Boosting HLT industry

Improvement of the **visibility** and **knowledge transfer** between academia and industry.

Support for **internationalization** and **commercialization** of the sector.

3. Public Administration as a driver of HLT Industry

Platforms for **natural language processing, speech technologies** and **automatic translation** in the public administrations.

Linguistic resources from **public administrations** according to the Directive on open data and the re-use of public sector information, also known as the '**Open Data Directive**' ([Directive \(EU\) 2019/1024](#))

4. Flagship Projects

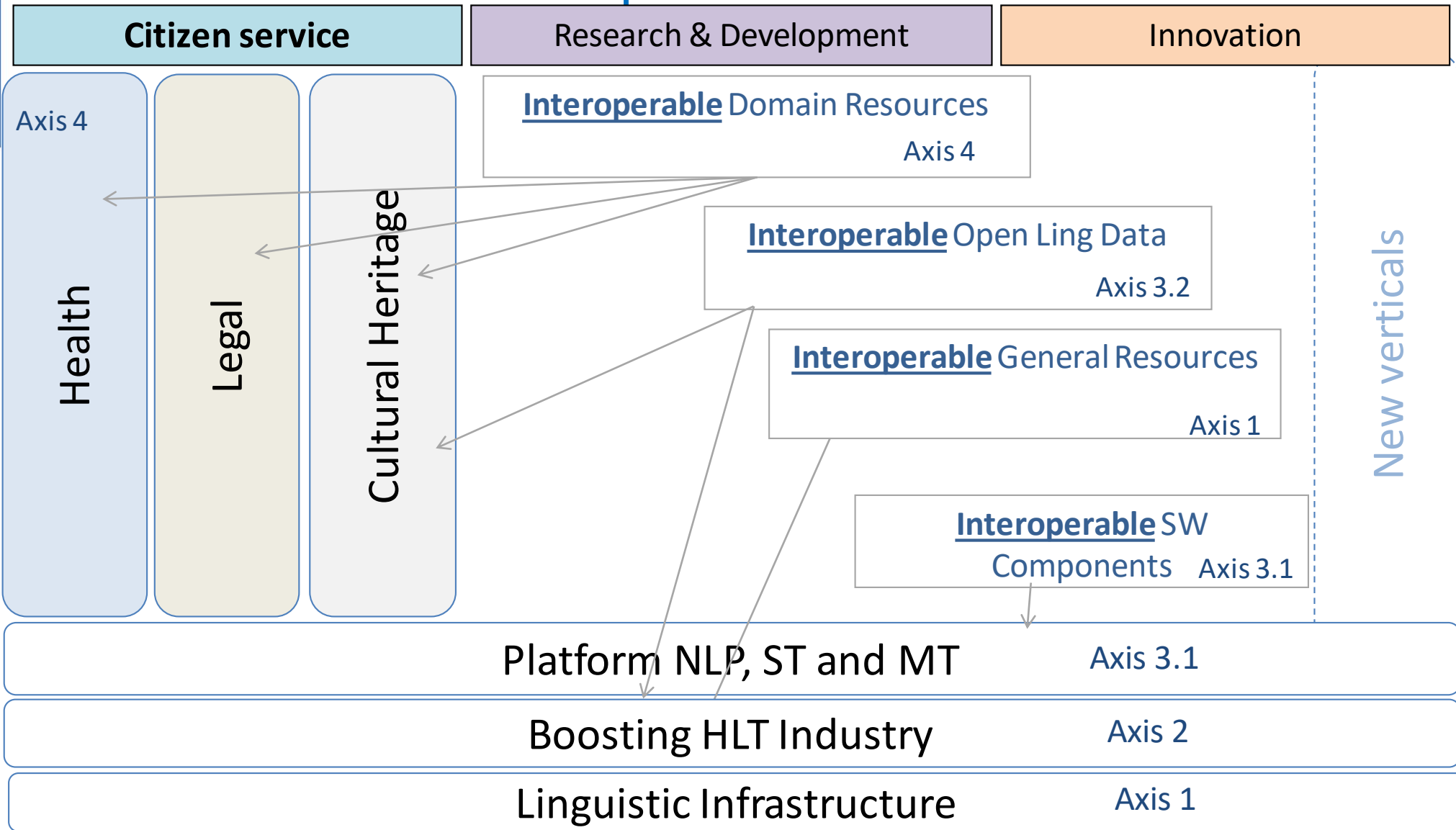
Health

Justice

Cultural Heritage

<http://www.plantl.es>

Plan TL: Roadmap



Axis 1. Language Infrastructures

- HLT infrastructure = **Resources** + **Processing**

Components/Tools

Interoperability between HLT resources and components is a **key** aspect for the **governance** and **sustainability** of the **HLT infrastructure**.

Achievements:

- RAE Corpus CAPITEL (Annotated Corpus 1000M words).
- News Agency EFE - SEAD agreement in process
- BSC processing of BNE Spanish web crawling in 10 years (5 TB)
- Legal Corpus (Spanish Legislations + Eurlex, etc.)



Shared Tasks- An instrument to ensure interoperability

- **Shared Tasks in Evaluation**

campaigns are considered as

Quality/Validation Instrument.

- Shared tasks as common practice in academia are competitive tasks where teams try to solve a problem given a dataset and a defined set of metrics.
- In industry, Kaggle is also an example of competitive tasks.

Axis 2: Boosting HLT Indus



- **SEAD Competitive Grants:** LT included in KETs Call 2019 (9M €).
- **Training & Capacity Building:** Training Modules & MOOCs are being developed for **INAP** public servant training for central, regional and local public administrations (MOOCs + training capsules)
- **Dissemination:** A number of Infodays on the different Flagship Projects: Competitive Intelligence, Health (BSC, Barcelona), Justice (Madrid-Dec 2019)

Axis 3: Public Administration as a Driver for HLT

Objective:

According to the Directive on open data and the re-use of public sector information, also known as the ‘**Open Data Directive**’ ([Directive \(EU\) 2019/1024](#)), there is a huge potential to use this data as resources for HLT..

Results

- **Automatic Translation Platform: EU CEF Projects: MTHub / ELRI / NTEU / NecData**
- ELRI Workshop in Spain (March 2019) at the European Commission headquarters in Madrid: Resource Infrastructure and Machine Translation Services for Public Institutions
- **Conversational Systems Platform**
- **Natural Language Processing Platform**
- Coordination with EU eInfra project OpeMinTed



Axis 4: Flagship Projects

Objective:

- **Innovating and improving existing public services for the citizens** using HLT.
- Facilitate the work of the Administrations in the **internal treatment of the information** and its use for defining and monitoring public policies.
- Generate interoperable and **reusable elements** for other projects.
- Immediate implementation of cross actions; using **general linguistic infrastructure** and **common platforms**.

Results- Health Sector

- Health **evaluation** campaigns
Recognition of substances in medical texts in Spanish
- Anonymization
- Medical **semantic indexing** in Spanish
CNIO: publication of developments in Github and Zenodo
- Corpus of Clinical Cases in Spanish: Morphosyntact

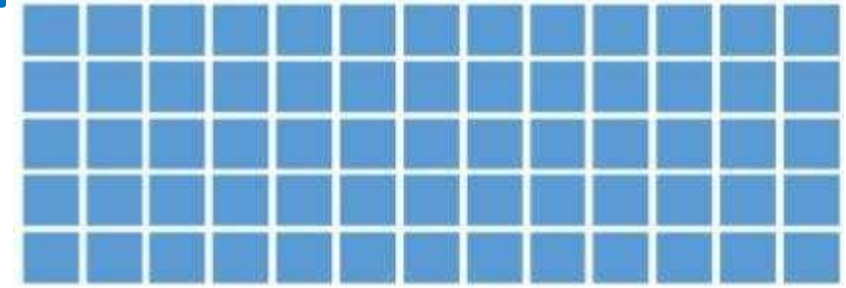
Guidelines

Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje



1. Why Language Technologies? -> Introduction
2. How is *Plan TL* structured? -> Framework of Plan TL
3. Where does *Semantic Interoperability of Metadata* lie within Plan TL?
4. What are the results of Plan? -> Use cases: Dialogue Systems, Machine Translation and Competitive Intelligence

3. Where does Semantic Interoperability of Metadata lie within Plan TL?



Structured Data

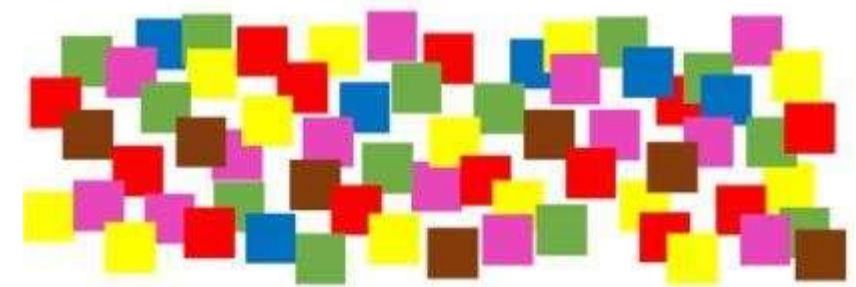
Our approach relies on two perspectives:

- **Structured Metadata**

→ Semantic Interoperability: Business Intelligence, ETL

- **Unstructured Data**

→ Semantics of content: NLP Pipelines, Topic Modeling, Machine Learning



Unstructured Data

3. Where does Semantic Interoperability of Metadata lie within Plan TL?

Interoperability is considered in the different phases:

- **Language Resources**
 - Metadata of corpora
 - Annotation Guidelines
- **Components and Tools**
 - Technology independent and open Access code



Guidelines

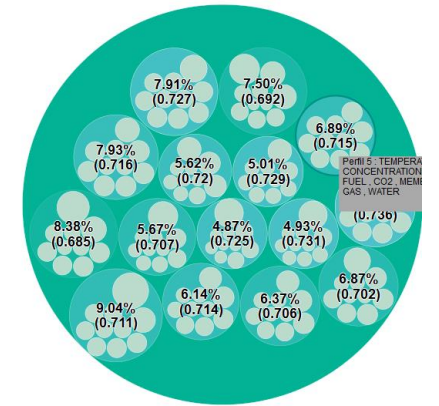
Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje



1. Why Language Technologies? -> Introduction
2. How is *Plan TL* structured? -> Framework of Plan TL
3. Where does *Semantic* Interoperability of *Metadata* lie within Plan TL?
4. What are the results of Plan? -> Use cases: Dialogue Systems, Machine Translation and Competitive Intelligence

4. Use cases: Competitive Intelligence NLP in RDI

- **Corpus Viewer** is an in house production service developed by the State Secretary for Digital Advancement (SEAD) within the framework of Plan TL.
- Corpus Viewer relies on Natural Language Processing (NLP), Machine Learning (ML) and Machine Translation (MT) to analyze structured metadata and unstructured textual data in large RDI corpora.
- The platform allows the decision maker and the policy implementer the possibility to analyze R&D&i information space (mainly patents, scientific publications and public grants) for evidence and knowledge-based policy making and implementation.



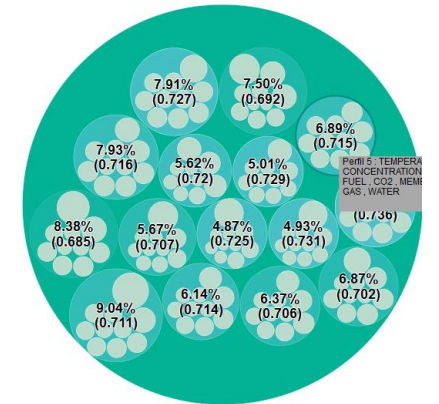
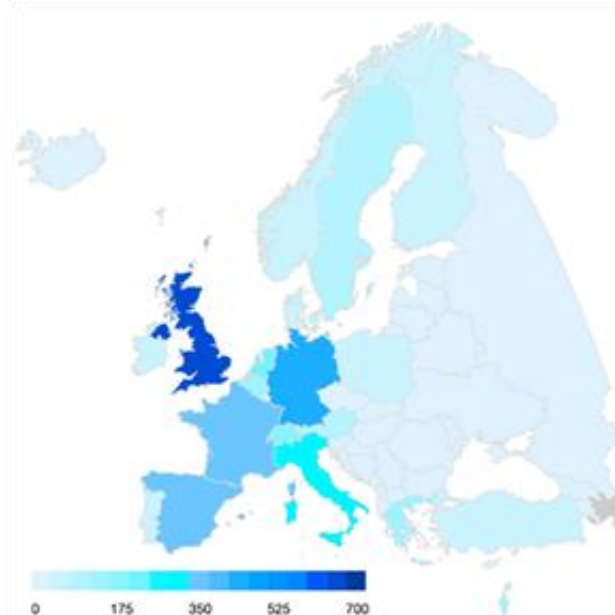
Corpus Viewer

4. Use cases: Competitive Intelligence NLP in RDI

Topic Analysis ¹

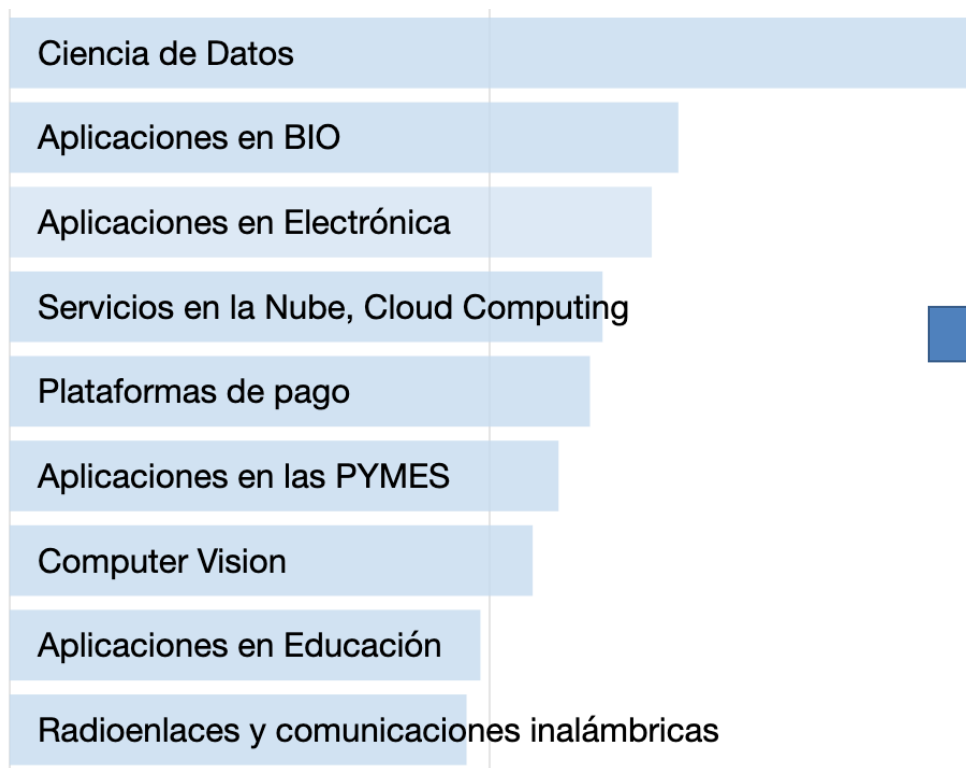


Programme ¹



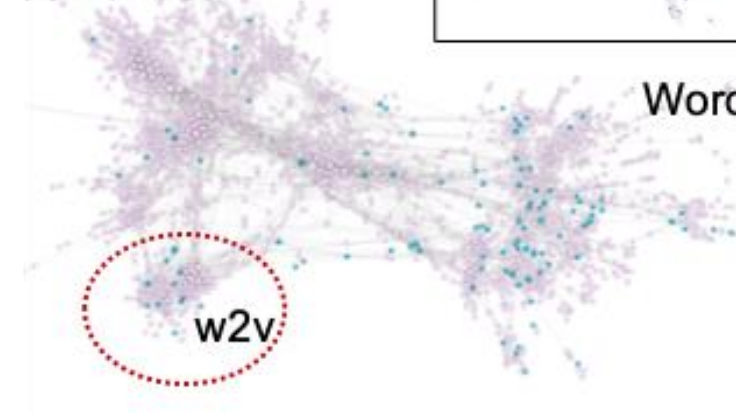
Corpus Viewer

4. Use cases: Competitive Intelligence NLP in RDI

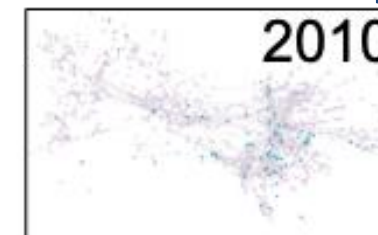


2018

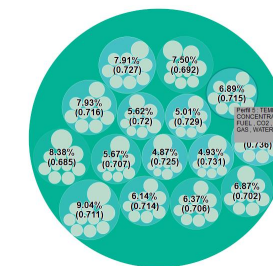
Latent semantics



2010



WordNet

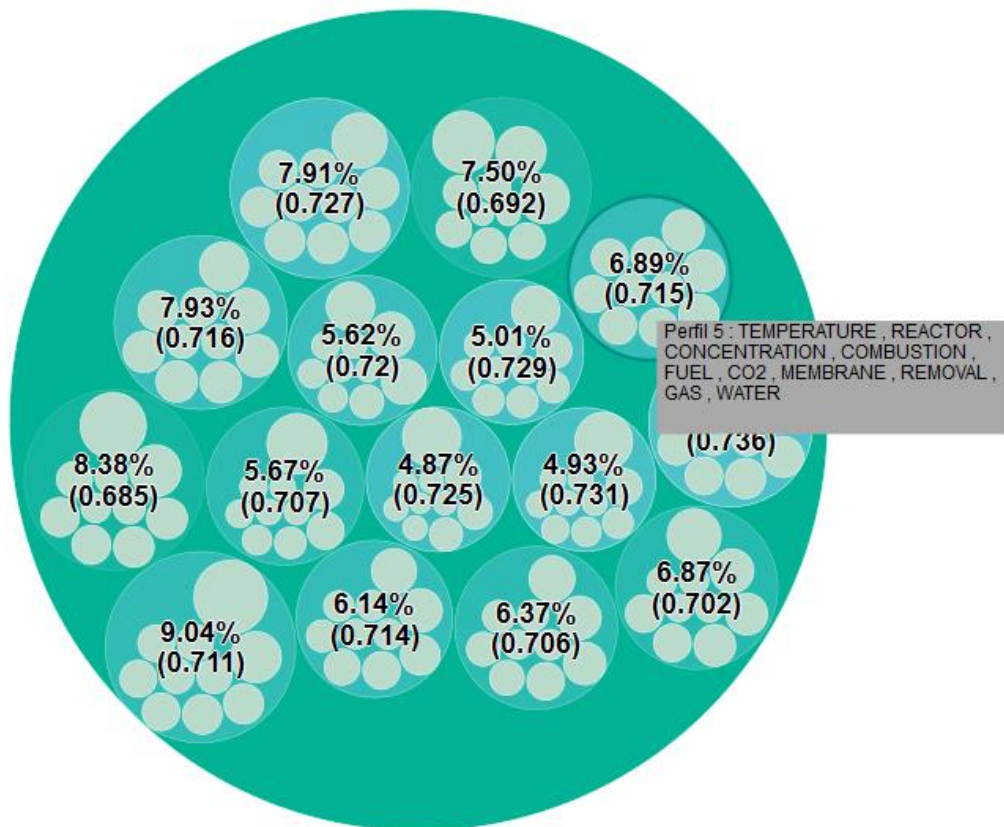


Corpus Viewer

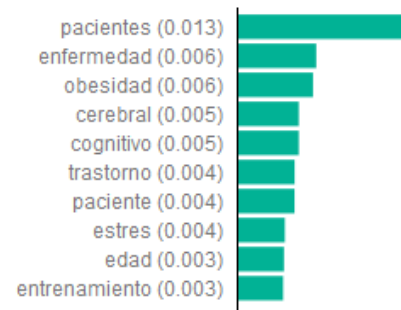
4. Use cases: Competitive Intelligence NLP in RDI

TOPICS OVERVIEW

Corpus: soopus Num. de documentos en el corpus: 13510 Algoritmo de perfilado: estatico Num. de perfiles: 15 Entropia media: 0 Fecha: 20/7/13/0 (5)

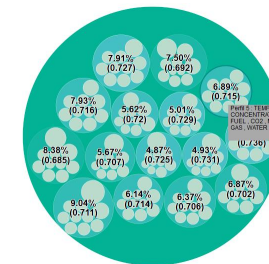


PALABRAS CARACTERÍSTICAS (RELEVANCIA)



DOCUMENTOS MEJOR CARACTERIZADOS POR ESTE PERFIL

%TOPIC	DOC ID	TITLE
100.00 %	PSI2012-35352	ALTERACIONES NEUROENDOCRINAS E INMUNITARIAS EN RATONES CON DIFERENTES ESTRATEGIAS DE AFRONTAMIENTO DEL ESTRES SOCIAL CRONICO. EFECTO DEL TRATAMIENTO CON UN ANTAGONISTA CRH1
100.00 %	PSI2008-00161	ESTUDIO DEL TRATAMIENTO CON UN AGONISTA SEROTONINERGICO DE ACCION RAPIDA SOBRE LOS EFECTOS CONDUCTUALE
100.00 %	PSI2011-24762	EFFECTOS DEL ESTRES SOCIAL EN EL CONDICIONAMIENTO DE LA PREFERENCIA DE LUGAR (CPL) INDUCIDO POR COCAINA Y MDMA (EXTASIS). IMPLICACION DEL SISTEMA DOPAMINERGICO Y GLUTAMATERGICO

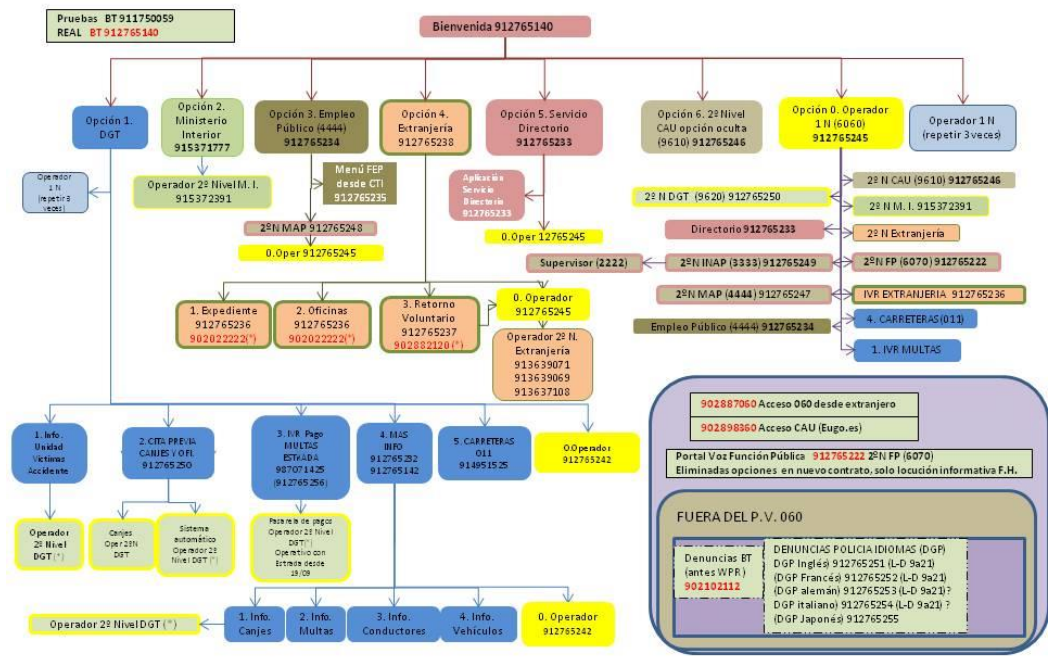


Corpus Viewer

- I. Introduction.
- II. IA. Machine Learning. NLP.
- III. What are Public Services doing?
- IV. **060 Service**
- V. Pilots. Chatbot –virtual assistant.
- VI. Work in progress.

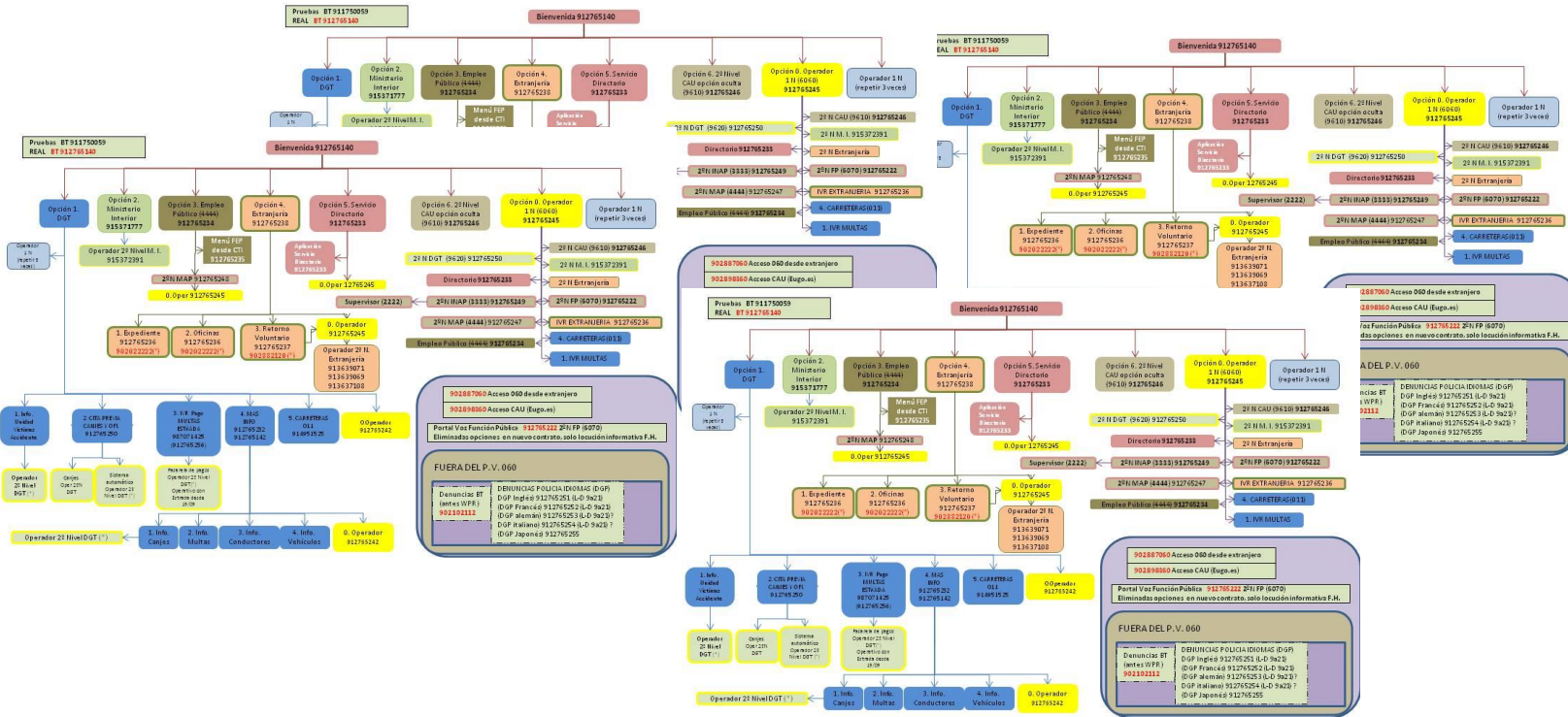
060 Service

- ✓ The 060 line is announced in 2006 with a certain objective, being the center point of attention for the Spanish citizens, in order to provide information regarding the Central Administration Services and to connect the different public services through one short number, 060.
- ✓ In the beginning, 060 provided information which was common for many Ministries.



060 Service

✓ What if many telephone information services are to merge into 060 system? How would we be capable of managing so many options?



060 Service

- ✓ How? An **open question** for the citizen to be answered.
- ✓ **NLP** at the service of making the Administration more efficient.
 - Intent detection.
 - Threshold definition to make decisions.
 - Training the open question system.
 - Threshold adapting.
 - “Self-learning”.

060 Service



Service

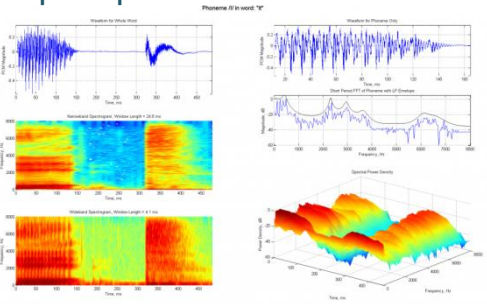
¿Detected?

Virtual Agent



Service detected

Open question



¿Over threshold defined?

Service detected



060

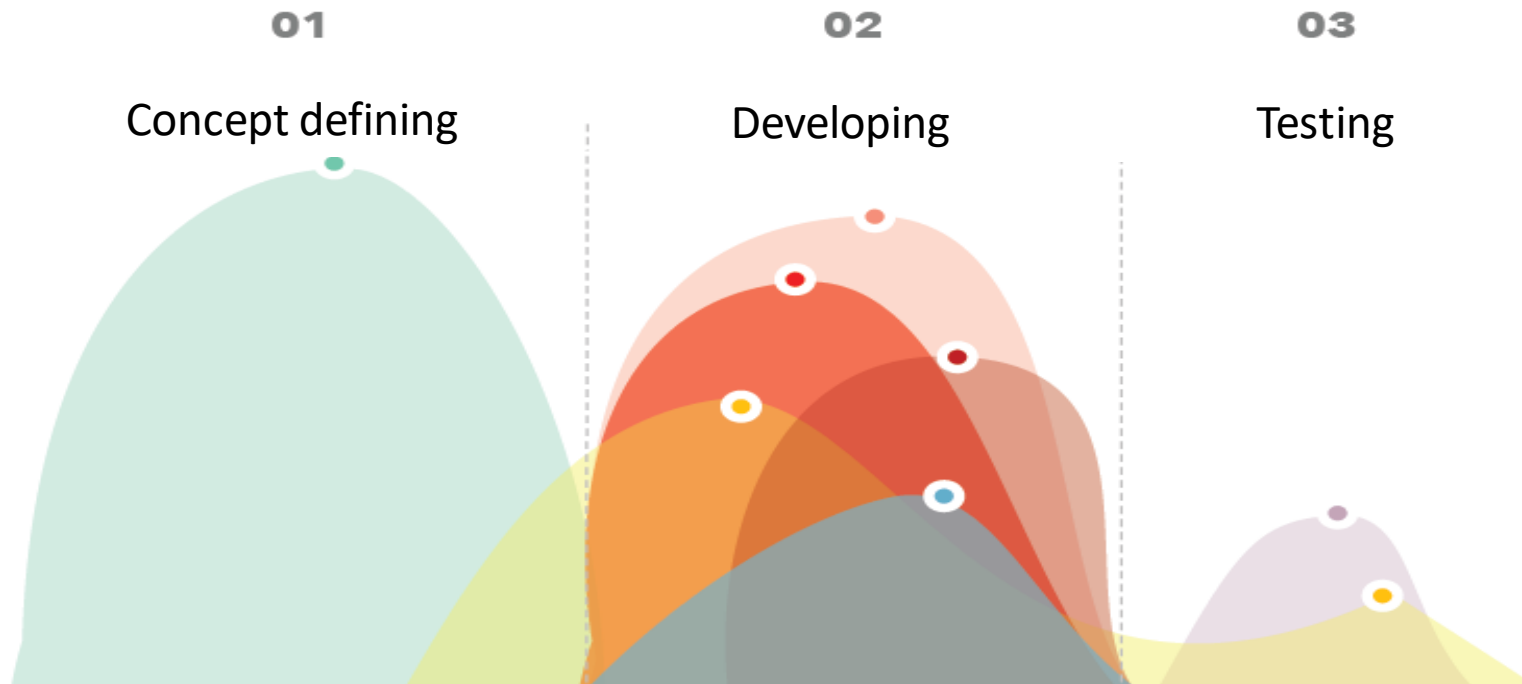
- I. Introduction.
- II. IA. Machine Learning. NLP.
- III. What are Public Services doing?
- IV. 060 Service
- V. **Pilots. Chatbot –virtual assistant.**
- VI. Work in progress.

Pilots. Chatbot –virtual assistant.

✓ **Conversational Assistant** for the citizen.

- FAQ.
- Transactional services as: fines, fees and tax payment, getting appointments, etc.
- Cross-channeling – developments made in text used in voice environments.
- Sentiment targeting
- Patterns detection
- Data analysis

Pilots. Chatbot –virtual assistant.



Pilots. Chatbot –virtual assistant.



Hola, soy el asistente virtual de la SGAD. Puedo ayudarle con dudas o problemas relacionados con Cl@ve PIN o Certificados Digitales. Puede probar con alguna de estas preguntas:

¿Qué es la Cl@ve PIN?

¿Cómo se obtiene una Cl@ve?

He olvidado mi contraseña

También puede escribir su propia consulta en el chat.

cómo obtengo una clave pin



¿Está usted dado de alta en el sistema Cl@ve?

Sí No No sé



Escriba su consulta aquí



Hola, soy el asistente virtual de la SGAD. Puedo ayudarle con dudas o problemas relacionados con Cl@ve PIN o Certificados Digitales. Puede probar con alguna de estas preguntas:

¿Qué es la Cl@ve PIN?

¿Cómo se obtiene una Cl@ve?

He olvidado mi contraseña

También puede escribir su propia consulta en el chat.

tengo problemas con mi clave pin



Indique qué clase de error está teniendo:

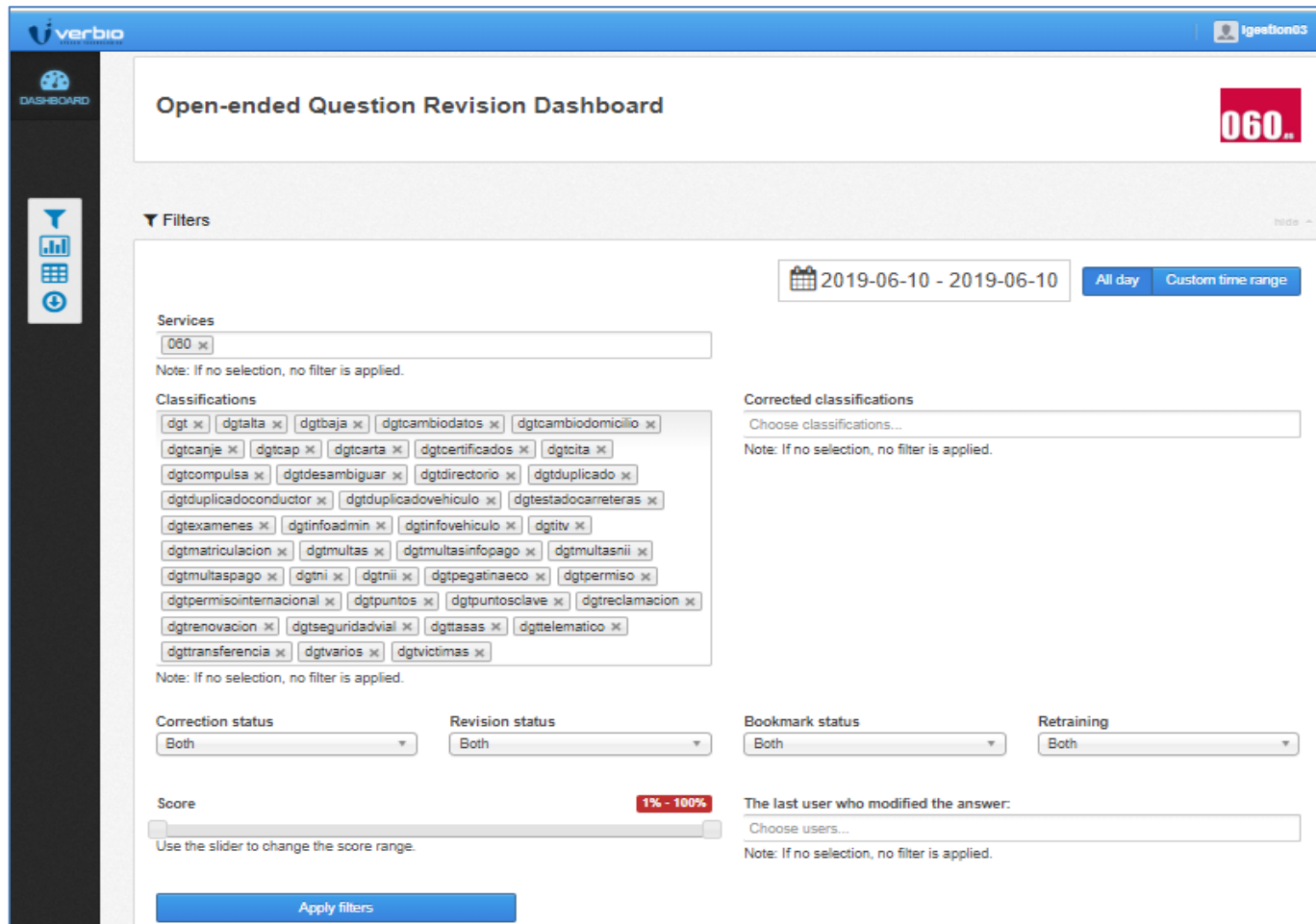
Datos de registro (email, DNI, etc.) Autenticacion SMS

DNI ya existe Error 66 No registrado de manera presencial



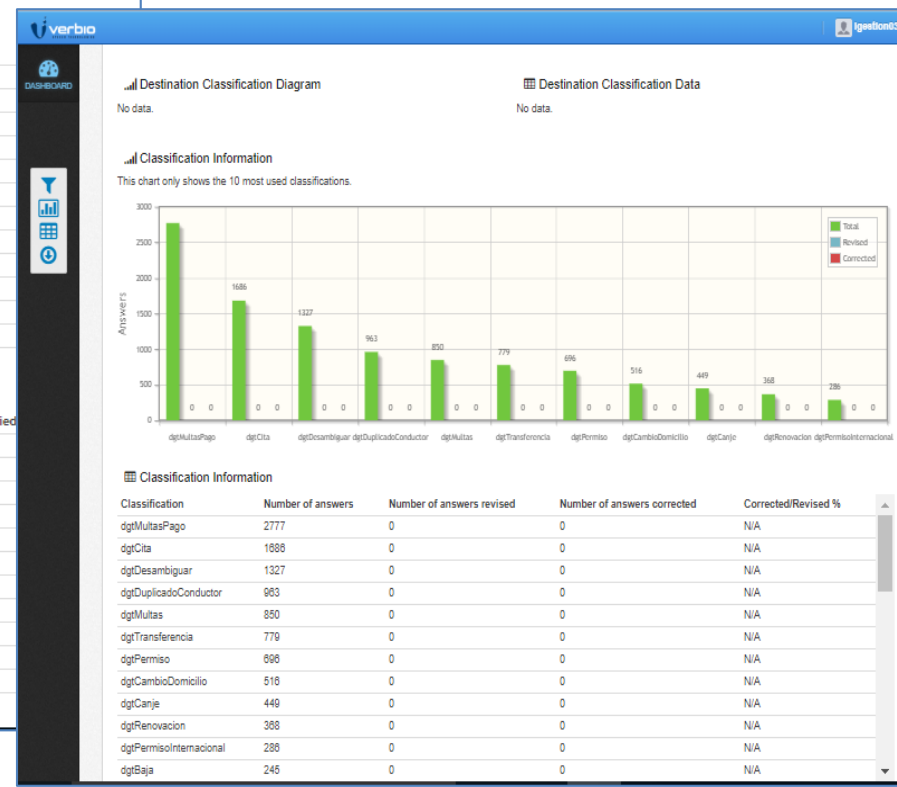
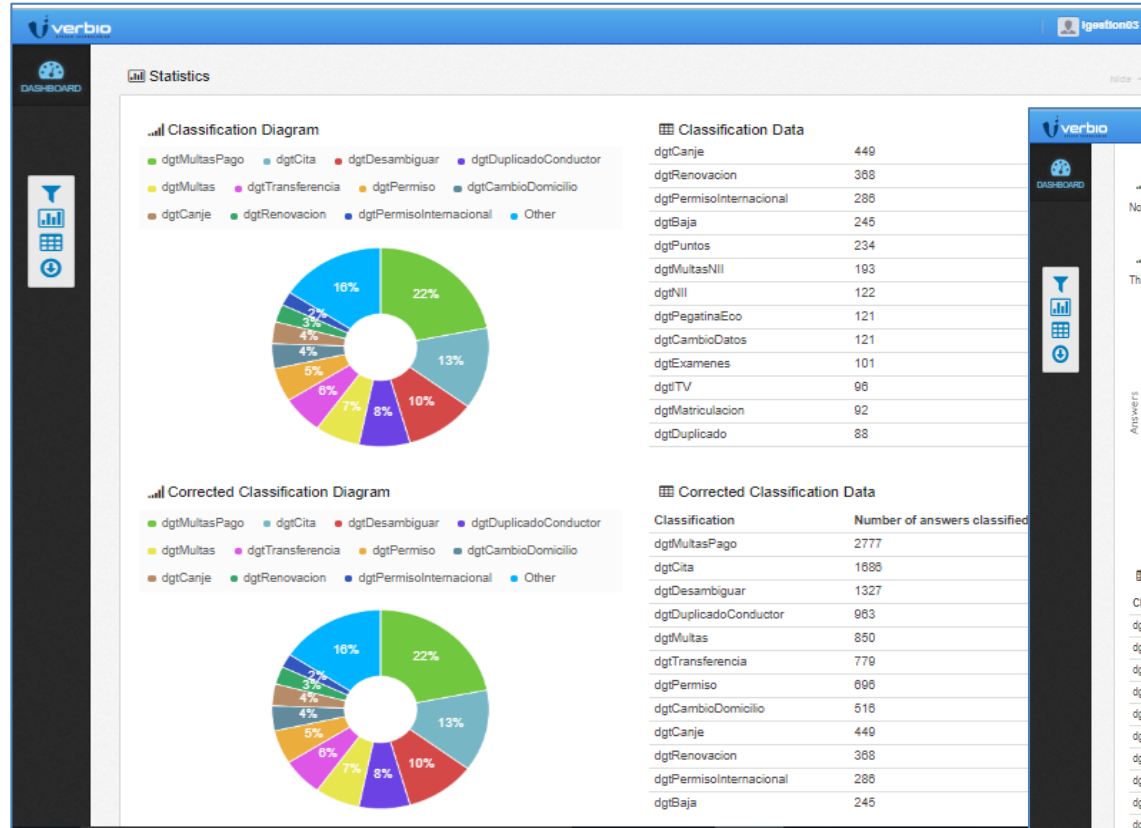
Escriba su consulta aquí

Pilots. Chatbot –virtual assistant.

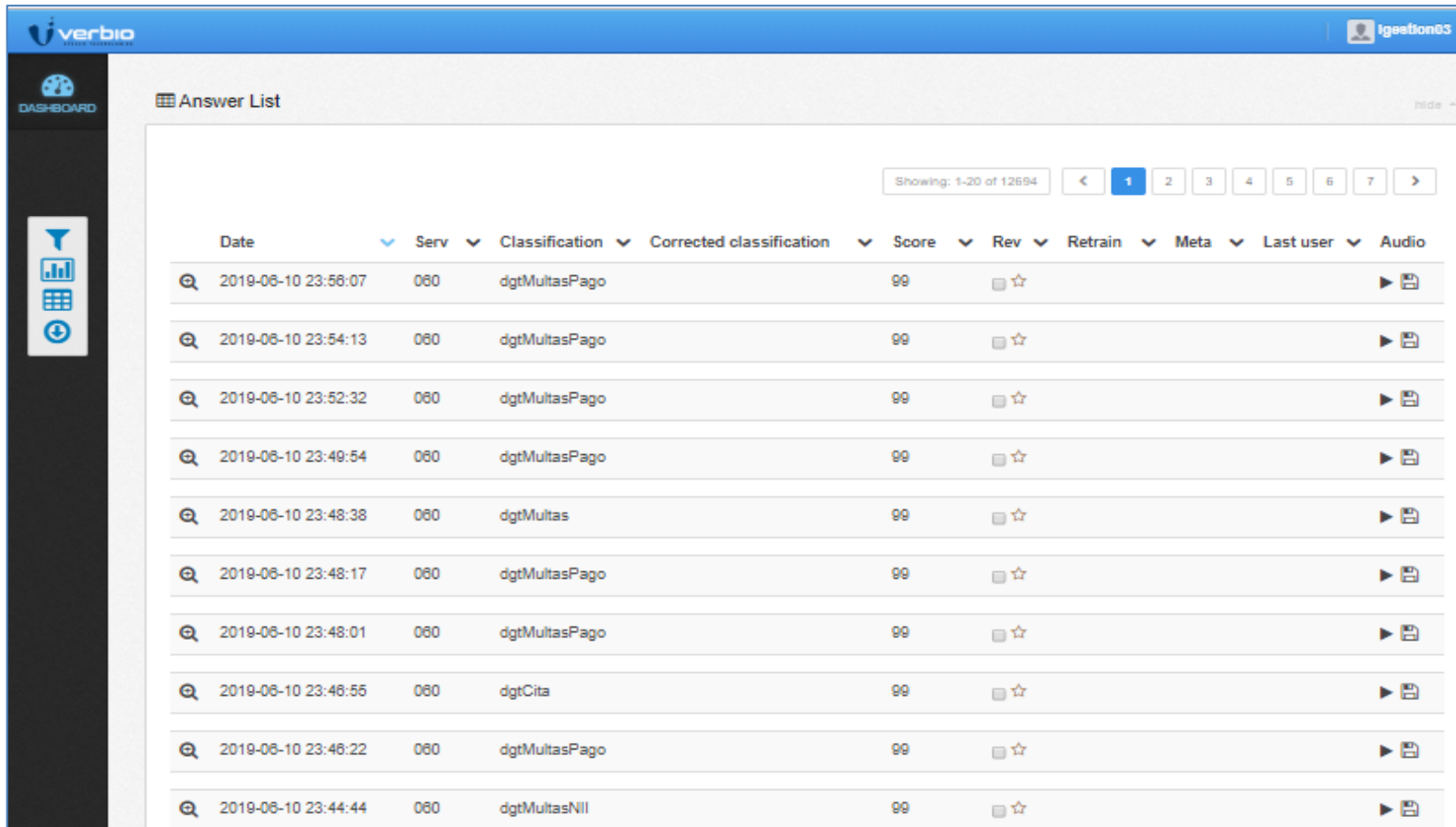


The screenshot shows the 'Open-ended Question Revision Dashboard' for the user 'lgeestlon63'. The dashboard includes a sidebar with navigation icons for 'DASHBOARD', a funnel, a bar chart, a calendar, and a refresh icon. The main content area features a date range filter set to '2019-06-10 - 2019-06-10' with 'All day' and 'Custom time range' options. Below this are several filter sections: 'Services' with a dropdown set to '060'; 'Classifications' with a grid of 30 tags such as 'dgt', 'dgtalta', 'dgtbaja', etc.; 'Correction status', 'Revision status', 'Bookmark status', and 'Retraining', each with a 'Both' dropdown; and a 'Score' slider set to '1% - 100%'. There are also input fields for 'Corrected classifications' and 'The last user who modified the answer'. An 'Apply filters' button is located at the bottom left of the main content area.

Pilots. Chatbot –virtual assistant.



Pilots. Chatbot –virtual assistant.



The screenshot shows the Verbio 'Answer List' interface. The top header includes the Verbio logo and the user 'Igeeflon03'. A sidebar on the left contains navigation icons for Dashboard, a funnel, a bar chart, a calendar, and a refresh button. The main content area displays a table of chatbot responses with the following columns: Date, Serv, Classification, Corrected classification, Score, Rev, Retrain, Meta, Last user, and Audio. The table shows 10 entries, all with a score of 99 and a date of 2019-08-10. The first entry is classified as 'dgtMultasPago', while the fifth is 'dgtMultas' and the seventh is 'dgtCita'. A pagination control at the top right indicates 'Showing: 1-20 of 12694' and shows page numbers 1 through 7.

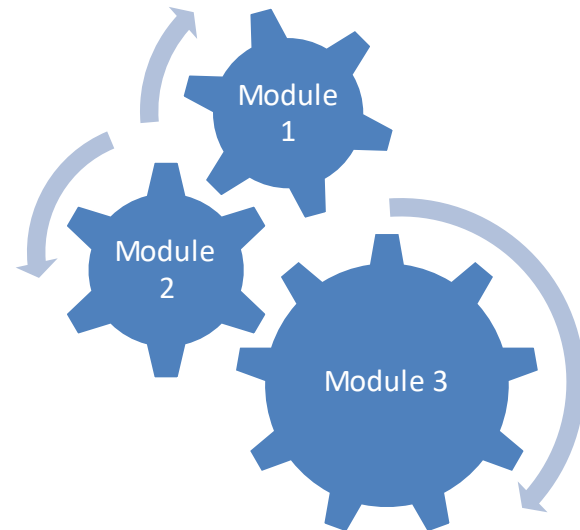
Date	Serv	Classification	Corrected classification	Score	Rev	Retrain	Meta	Last user	Audio
2019-08-10 23:58:07	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:54:13	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:52:32	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:49:54	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:48:38	060	dgtMultas		99	☐ ☆				▶ 📄
2019-08-10 23:48:17	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:48:01	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:46:55	060	dgtCita		99	☐ ☆				▶ 📄
2019-08-10 23:46:22	060	dgtMultasPago		99	☐ ☆				▶ 📄
2019-08-10 23:44:44	060	dgtMultasNil		99	☐ ☆				▶ 📄

- I. Introduction.
- II. IA. Machine Learning. NLP.
- III. What are Public Services doing?
- IV. 060 Service.
- V. Pilots. Chatbot –virtual assistant.
- VI. **Work in progress.**

Work in progress.

✓ **Under the Natural Language Plan:**

- Common platform
- Semantic Interoperability
- Module Interoperability
- Common corpus
- Re-use



Future.

✓ Future:

- Designing of plans and strategies
- Data driven smart government
- Common corpus
- Re-use



Thank you

salvador.estevan@correo.gob.es

