

Lattice-Based conceptualization (DIMH project)

Ana García-Serrano
ETSI Infomática, UNED

(In collaboration with Ángel Castellanos and Juan Cigarrán)

Catálogo de Mapas, Planos y Dibujos del Archivo General de Simancas

- Permite descargar datos estructurados (XML, RDF DC).



The screenshot shows a Firefox browser window displaying the website 'Catálogo Colectivo de la Red de Bibliotecas de los Archivos Estatales'. The address bar shows the URL 'www.mcu.es/ccbae/es/mapas/principal.cmd'. The page features a header with the Spanish flag, the text 'GOBIERNO DE ESPAÑA', 'MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE', and 'SECRETARÍA DE ESTADO DE CULTURA'. The main navigation menu includes 'Presentación', 'Búsqueda', 'Búsqueda avanzada', 'Hemeroteca', 'Mapas, Planos y Dibujos', 'Estadísticas', 'Contenido RSS', and 'Repositorio OAI'. The 'Mapas, Planos y Dibujos' section is active, showing a search bar with 'Navegación', 'Busqueda Simple', and 'Busqueda Avanzada' options. Below the search bar, the title 'Catálogo Colectivo de las Colecciones de Mapas, Planos y Dibujos de los Archivos Estatales' is displayed, followed by the subtitle 'Colección de Mapas, Planos y Dibujos del Archivo General de Simancas'. Four thumbnail images are shown, each with a caption: 'Material cartográfico AGS', 'Material gráfico AGS', 'Objetos tridimensionales AGS', and 'Material textual AGS'.

http://www.mcu.es/ccbae/es/consulta/resultados_busqueda.cmd?tipo_busqueda=mapas_planos_dibujos&posicion=1&id=30485

Tarea: Analizar el corpus

| | |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sección: | Material gráfico AGS |
| Número de control: | BAB20100018941 |
| Autor: | Huet, Luis  |
| Título: | Plano y perfiles que manifiestan el estado en que se alla la Real obra de Fuerte-Príncipe en 3 de 1779 [Material gráfico no proyectable] / [rúbrica] Luis Huet |
| Área de datos: | Escala [ca.1:816], 200 varas reales [=20,6 cm] |
| Publicación: | Habana, 30 de junio de 1779 |
| Descripción física: | 1 plano : ms., col. ; 36,5 x 47 cm |
| Notas: | Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : p. 576 Tinta y colores a la aguada ocre y encarnado. Explicación con clave alfabética Manuscrito sobre papel. AGS. Secretaría de Guerra, Legajos, 03222. Acompaña a carta y relación delas obras de don al Conde de Ricla de 1 de julio y de la misma fecha del plano |
| Materia / geográfico: | Fortificaciones-La Habana-Dibujos  La Habana-Edificios, estructuras, etc.-Dibujos  |
| Género / forma: | Dibujos de arquitectura-España-S.XVIII  |
| Nombre jerárquico lugar: | Cuba La Habana  |

Ficha: 176927

Ficha: 176927 (RDFDC)

<rdf:Description>

<dc:relation>Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : p. 570

</dc:relation>

<dc:coverage>La Habana</dc:coverage>

...

<dc:coverage>-S.XVIII</dc:coverage>

<dc:title>Plano y perfiles que manifiestan el estado en que se alla la Real obra de Fuerte de San Carlos de Príncipe en 30 de Junio de 1779 [Material gráfico no proyectable]</dc:title>

<dc:description>AGS. Secretaría de Guerra, Legajos, 03222. Acompaña a carta y relación de las obras de don Luis Huet al Conde de Ricla de 1 de julio y de la misma fecha del plano</dc:description>

<dc:description>Tinta y colores a la aguada ocre y encarnado. Explicación con clave alfabética</dc:description>

<dc:description>Manuscrito sobre papel.</dc:description>

<dc:type>Ilustraciones y Fotos</dc:type>

<dc:language>spa</dc:language>

<dc:date>1779</dc:date>

<dc:creator>Huet, Luis</dc:creator>

Primera aproximación (abril 2014)

- Tarea 1:
 - ✓ Estudiar el corpus (7792 fichas) correspondientes a mapas, planos y dibujos del Archivo General de Simancas (AGS).
 - ✓ Construir la versión XML enriquecida (NER, POS, Lemas)
- Tarea 2: Indexación
 - ✓ Búsqueda automática
 - ✓ Búsqueda configurable

Tarea: Enriquecer lingüísticamente el corpus (no supervisado)

<Ficha id="176927">
<Tematica>La Habana-Edificios, estructuras, etc.-Dibujos</Tematica>
<Materia>-La Habana</Materia>
<Materia>-España</Materia>
<Formato>image/jpeg</Formato>
<Idioma>spa</Idioma>
<Materia>La Habana</Materia>
<Materia>-S.XVIII</Materia>
<Notas>AGS. Secretaría de Guerra, Legajos, 03222. Acompaña a carta y relación delas obras de don Luis Huet al Conde de Ricla de 1 de julio y de la misma fecha del plano</Notas>
<Creador>Huet, Luis</Creador>
<Publicacion>1779</Publicacion>
<Notas>Manuscrito sobre papel.</Notas>
<Materia>Cuba-La Habana</Materia>
... </Ficha>

<nes>La_Habana Fortificaciones-Dibujos_Dibujos Dibujos Luis_Huet Huet AGS Cuba España Real_Sociedad Real Conde_de_Ricla Fuerte-Príncipe Luis_La_Habana </nes>
<nes_person>Luis_Huet Huet Fuerte-Príncipe</nes_person>
<nes_organization>Real_Sociedad Real</nes_organization>
<nes_location>La_Habana Cuba España </nes_location>
<nes_misc>Fortificaciones-Dibujos_Dibujos Dibujos AGS Conde_de_Ricla Luis_La_Habana</nes_misc>
<lemas> plano y perfil que manifestar el estado en que se allá el Real obrar de Fuerte - Príncipe en 30 de Junio de 1779 ... </lemas>
<sintagmas> plano_perfil real fuerte_principe junio_1779_material_grafico proyectable_huet luis_habana_edificio dibujos **Ficha: 176927**
... </Ficha>

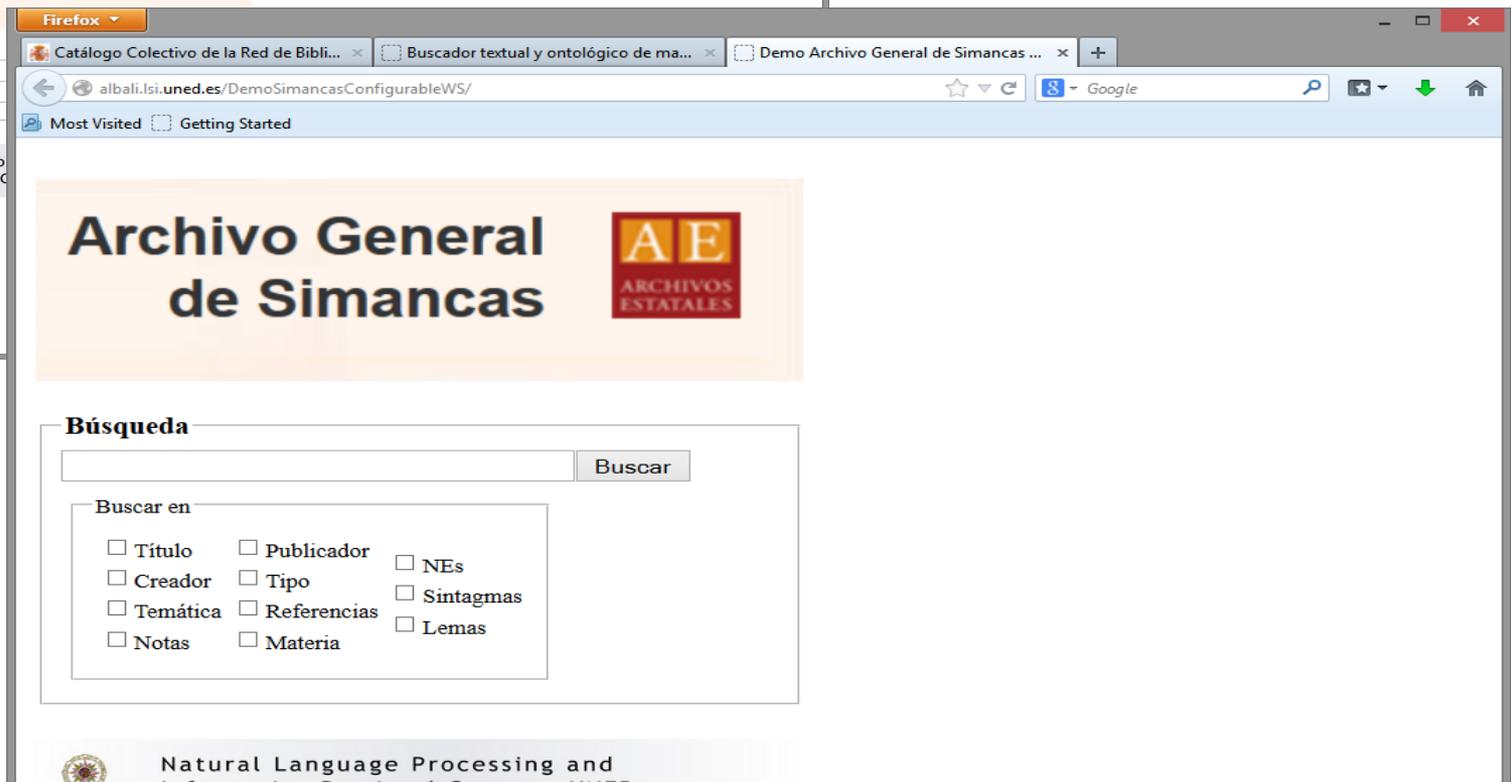
Tarea: Herramienta de búsqueda

Búsqueda Automática: <http://albali.lsi.uned.es/DemoSimancasWS/>



Búsqueda Configurable:

<http://albali.lsi.uned.es/DemoSimancasConfigurableWS/>



Interacción con los historiadores (1)

- Vocabulario

| | | | |
|------------------------------------------|------------------------------------|------------------------------------------------------------------|----------------------------------------------------|
| Autores (nombres citados) | Cramer. Agustín | Patiño, José | Buceta, Ventura |
| | | | Cafaro, Donato |
| | Martínez de la Vega, Dionisio | Próspero, Jorge (Marqués de Verboom) [1665-1744] | Antonio Picchiatti, |
| | | Montero de Espinosa, Jose(ph) (capitán de navío) | Francesco Antonio |
| | Caballero (Cavallo), Bruno | Guemes y Horcasitas, Juan Francisco (Gobernador de La Habana) | Castellón, José (Joseph) Domingo y Cueba, |
| | Liguera, Juan de Reggio, Andrés | Marqués de la Ensenada Tavira, Ramón | Francisco |



Interacción con los historiadores (1)

- Vocabulario terminológico

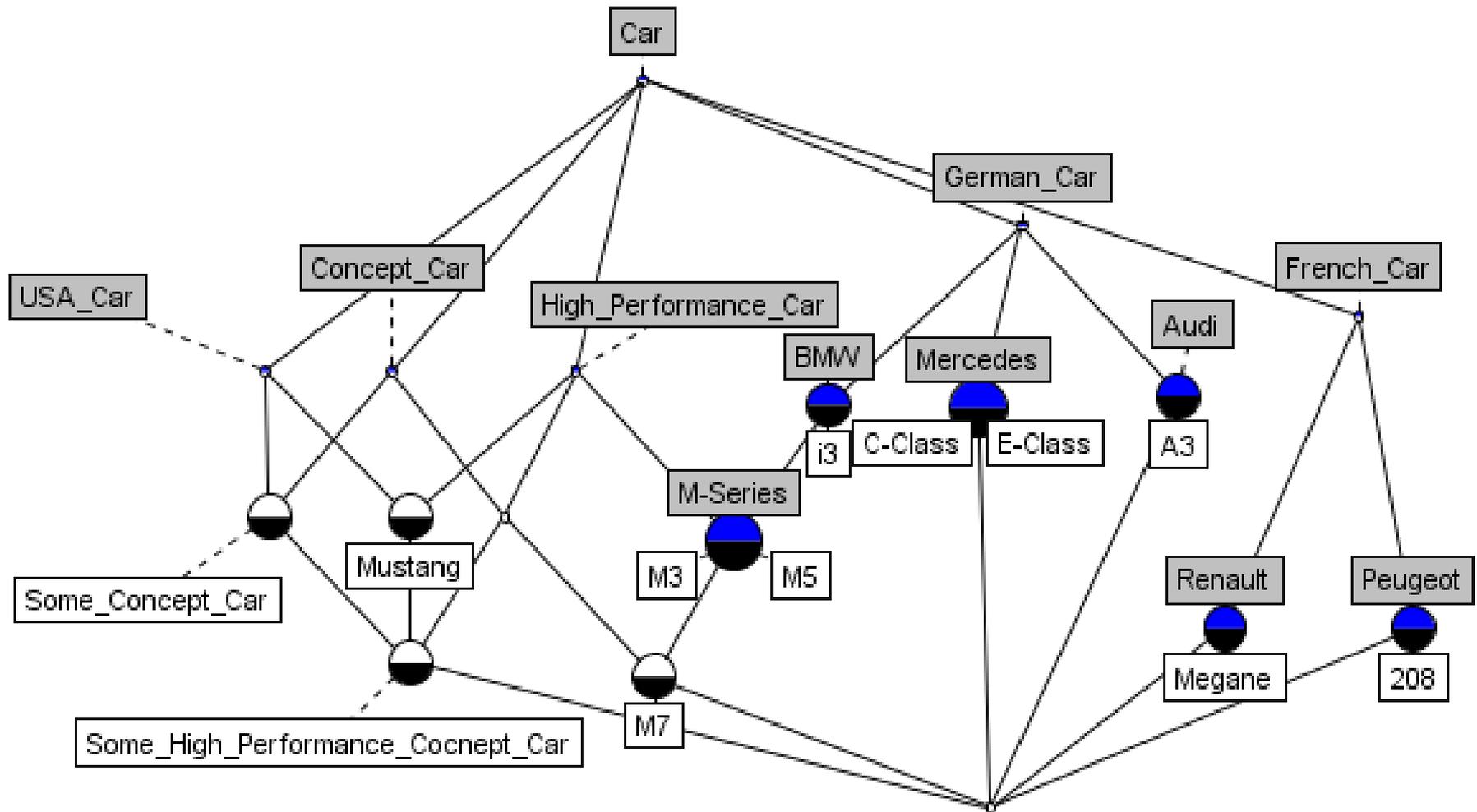
Medidas

| | |
|--------------------------------|-------------------|
| toesas/tuesas/tuesas reales | codos |
| escala | baras |
| pies reales | braseaje/braceaje |
| varas castellanas/ de Castilla | canas |
| varas reales | pies geométricos |
| varas cúbicas | pies comunes |
| pies castellanos/de Castilla | pie |
| pulgada | milla |

FCA for Content Organization

- A mathematical theory (Formal Concept Analysis, **FCA**) (Development of our own **framework**)
- **Formal Concept Analysis ¿a suitable solution?**
 - In general: it allows the organization of objects according to their shared attributes into a generalization-specification relationship.
 - Concrete:
 - organization of latent structure according to the shared terms in thematic-based concepts.
 - Lattice-based organization (visualization, navigable structure)
 - Datasets: Dbpedia, RepLab2013

Dbpedia: Domain-based FCA Modelling



Representation through a **Concept-based Knowledge Organization**

Formal Concept Analysis [Wille, 1992]

RepLab2013: Filtering data extracted from LOD



sheeba
@sheebapkv

M3 & M5 M Performance Editions - Web Exclusive:
M enthusiasts in the UK will be rejoicing at the
announc...

LOD

LOD_FCA

M3: BMW, German_Car, M-Series,
Car, High_Performance_Car...

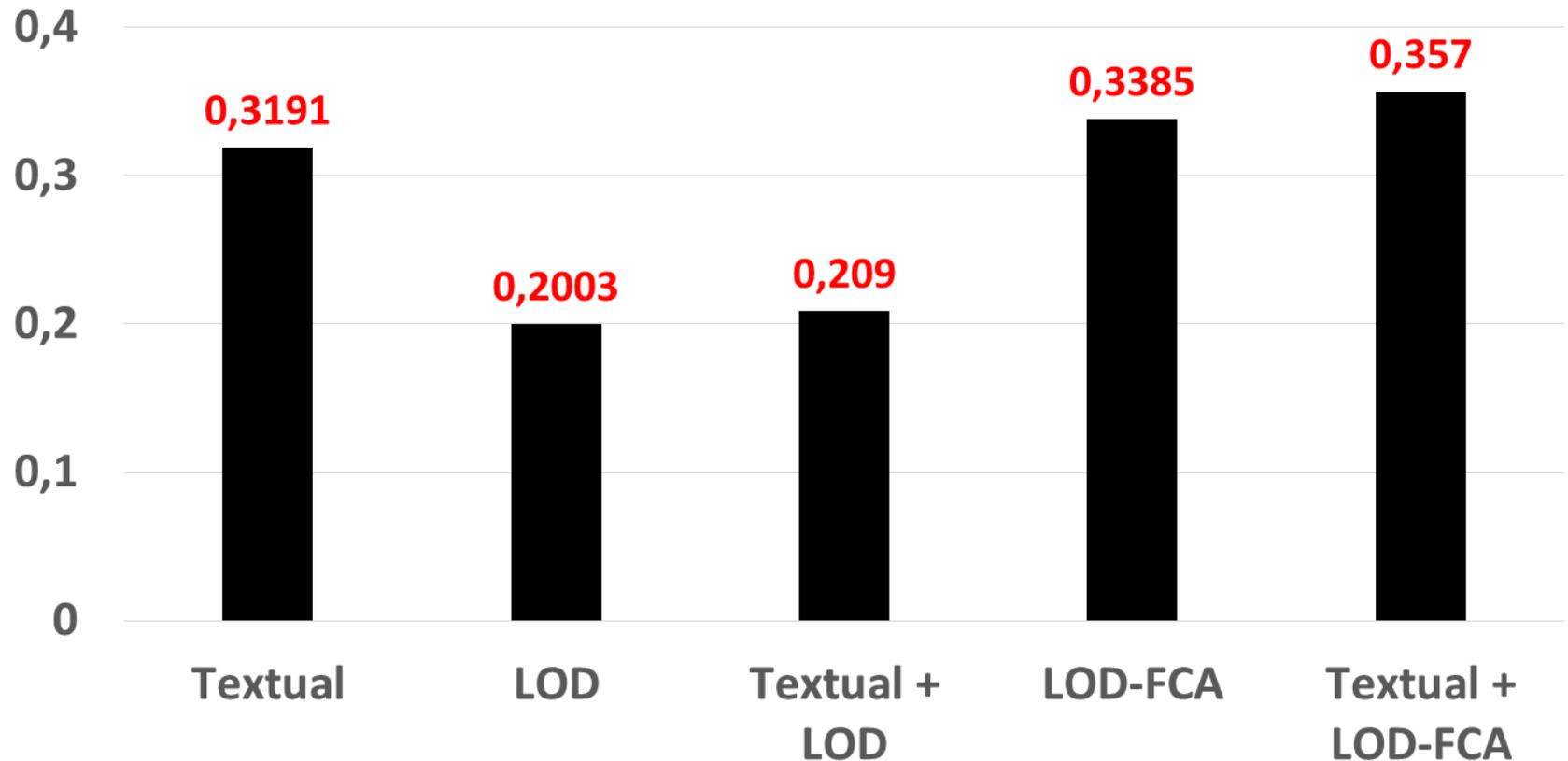
M3: BMW, M-Series, High_Performance_Car

M5: BMW, German_Car, M-Series,
Car, High_Performance_Car...

M5: BMW, M-Series, High_Performance_Car

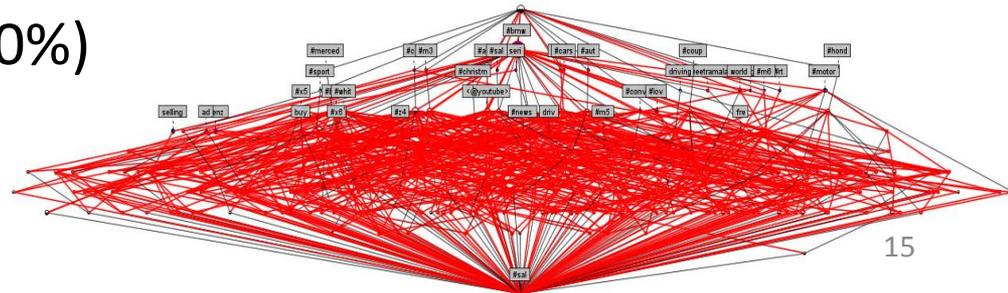
Results: Replab 2013 Topic Detection

F-measure



RepLab2013 Experiments

- Real settings: Huge lattices!!!: **how to select the best attributes, reducing the noise without losing knowledge.**
- Pre-selection of the top representative features
 - only the attributes that mostly appears in the generated formal concepts.
 - Lower & Upper Threshold
 - ¿how to test when no evaluation data available?
 - Currently: (1% and 50%)



Applied Research: Own ToolKit

- Modelling step
 - Library for KLD Pre- Filtering
 - FCA Algorithm (Jbrained, la4j for huge matrix)
- Topic detection and Recommendation
 - Graph Representation (Neo4j, graph DB based on Lucene)
 - Recommendation Algorithms (Code for graph navigation)
- Interface to facilitate the user interaction with the Formal Concept Lattice
 - (Alpha version using Google Web Toolkit)
- Tools for Lab experimentation (refinement, evaluation)

la4j



DIMH: FCA-Based Modelling

- FCA components:
 - Objects = DIMH Files
 - Attributes = terms + special features
 - Formal Concepts = similar DIMH files together according to their features
 - Structure = Formal concepts ordered in a lattice
 - Navigable Structure
 - Informative Representation

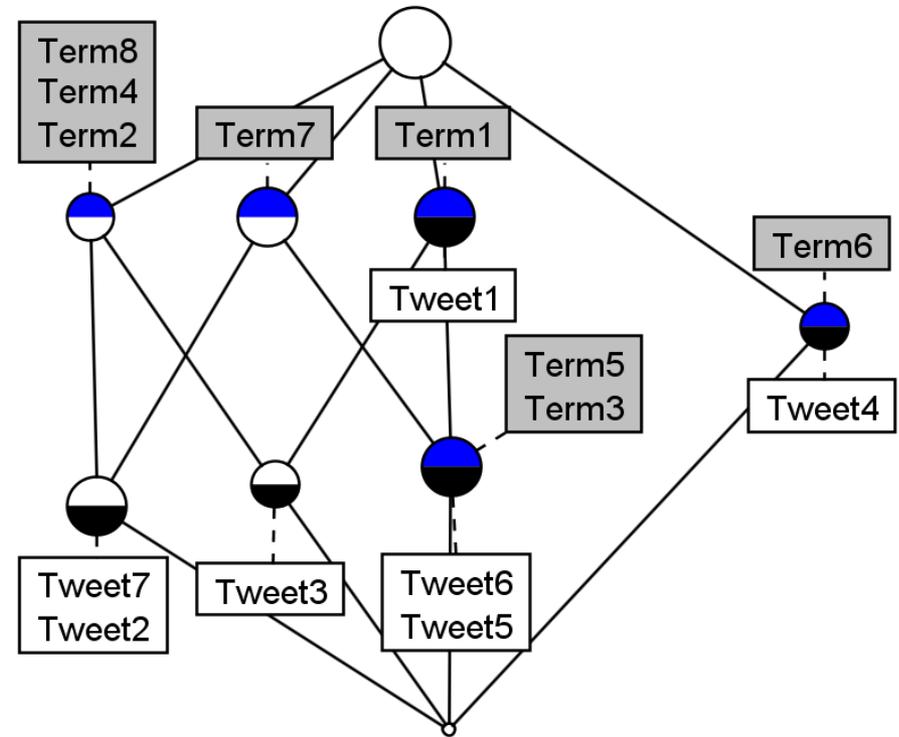
Formal Context Generation

| | |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sección: | Material gráfico AGS |
| Número de control: | BAB20100018941 |
| Autor: | Huet, Luis  |
| Título: | Plano y perfiles que manifiestan el estado en que se alla la Real obra de Fuerte-Príncipe en 30 de Junio de 1779 [Material gráfico no proyectable] / [rúbrica] Luis Huet |
| Área de datos: | Escala [ca.1:816], 200 varas reales [=20,6 cm] |
| Publicación: | Habana, 30 de junio de 1779 |
| Descripción física: | 1 plano : ms., col. ; 36,5 x 47 cm |
| Notas: | Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : p. 576 Tinta y colores a la aguada ocre y encarnado. Explicación con clave alfabética Manuscrito sobre papel. AGS. Secretaría de Guerra, Legajos, 03222. Acompaña a carta y relación delas obras de don Luis Huet al Conde de Ricla de 1 de julio y de la misma fecha del plano |
| Materia / geográfico: | Fortificaciones-La Habana-Dibujos  La Habana-Edificios, estructuras, etc. Dibujos  |
| Género / forma: | Dibujos de arquitectura |
| Nombre jerárquico lugar: | Cuba-La Habana  |
| Autores secundarios: | Huet, Luis  |
| Tipo de publicación: |  Ilustraciones y F |
| Ejemplares: | Archivo General de S Préstamo:  Dispon  Copia Digital |

| | Term ₁ | Term ₂ | ... | Term _n |
|--------------------|-------------------|-------------------|-----|-------------------|
| Dfile ₁ | X | | ... | X |
| ... | ... | ... | ... | ... |
| Dfile _n | | X | ... | X |

Concept Lattice Generation

| | Term ₁ | Term ₂ | Term ₃ | Term ₄ | Term ₅ | Term ₆ | Term ₇ | Term ₈ |
|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Dfile ₁ | X | | | | | | | |
| Dfile ₂ | | X | | X | | | | |
| Dfile ₃ | X | X | | X | | | | |
| Dfile ₄ | | | | | | | | |
| Dfile ₅ | X | | X | | | | | |
| Dfile ₆ | X | | X | | | | | |
| Dfile ₇ | | X | | X | | | | |



Attribute Selection: KLD pre-filter (1)

- The attributes with a **percentage lower than the lower threshold** are removed and the attributes with a **percentage higher than the upper** are automatically taken.

| | Fichas | Atributos | Relaciones | Conceptos Formales |
|------------|---------------|------------------|-------------------|-------------------------------|
| ALL | 7792 | 103 | 29874 | 17501 |

Attribute Selection: KLD pre-filter (2)

- **(Some) Terminology and NE (defined by experts)**
- Simple Available Vocabulary of the Domain

| | Fichas | Atributos | Relaciones | Conceptos Formales |
|------------|---------------|------------------|-------------------|-------------------------------|
| ALL | 7792 | 36 | 13719 | 1197 |
| | | | | 17501 |

Interface para navegar

| |
|----------------------------------|
| [proyectable, sxvi] (7) |
| [sxvi, negra] (7) |
| [proyectable, sxvi, negra] (3) |
| [proyectable, españa, negra] (3) |
| [españa, sxvi, negra] (3) |
| [españa, sxvi] (13) |
| [españa, negra] (14) |
| [proyectable, negra] (16) |
| [proyectable, españa] (25) |
| [sxvi] (77) |
| [negra] (89) |
| [proyectable] (188) |
| [españa] (266) |

Showing information about the concept: [españa, sxvi, negra] (3)

Hanse de hazer para la isla de Canarias quatro culebrinas como la que aquí va designada... [Material gráfico no proyectable]

1587

Idioma: spa

-España -S.XVI AGS. Guerra y Marina, Legajos, 00199, 152. Con carta de don Juan de Acuña al Rey. Lisboa, 25 de julio de 1587. Tinta negra. explicación Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : 412. Hanse de hazer para la isla de Canarias quatro culebrinas como la que aquí va designada... [Material gráfico no proyectable] Canarias España Juan_de_Acuña Lisboa Lisboa Fotos_Referencias Juan_de_Dibujos España Lisboa Lisboa Fotos_Referencias Juan_de_Dibujos Juan_de_Acuña

Plano de una grada para botar navíos [Material gráfico no proyectable]

1594

Idioma: spa

-España -S.XVI Manuscrito sobre papel Tinta negra. Con explicación AGS. Estado, Legajos, 00170. Remitido por el capitán Ricardo Butiler en a S.M., Madrid 9 de octubre de 1594. Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : p. 549 Plano de una grada para botar navíos [Material gráfico no proyectable] AGS Astilleros-Dibujos_Manuscrito España Estado Madrid Madrid capitán_Ricardo_Butiler España Madrid Madrid capitán_Ricardo_Butiler Astilleros-Dibujos_Manuscrito Estado capitán_Ricardo_Butiler

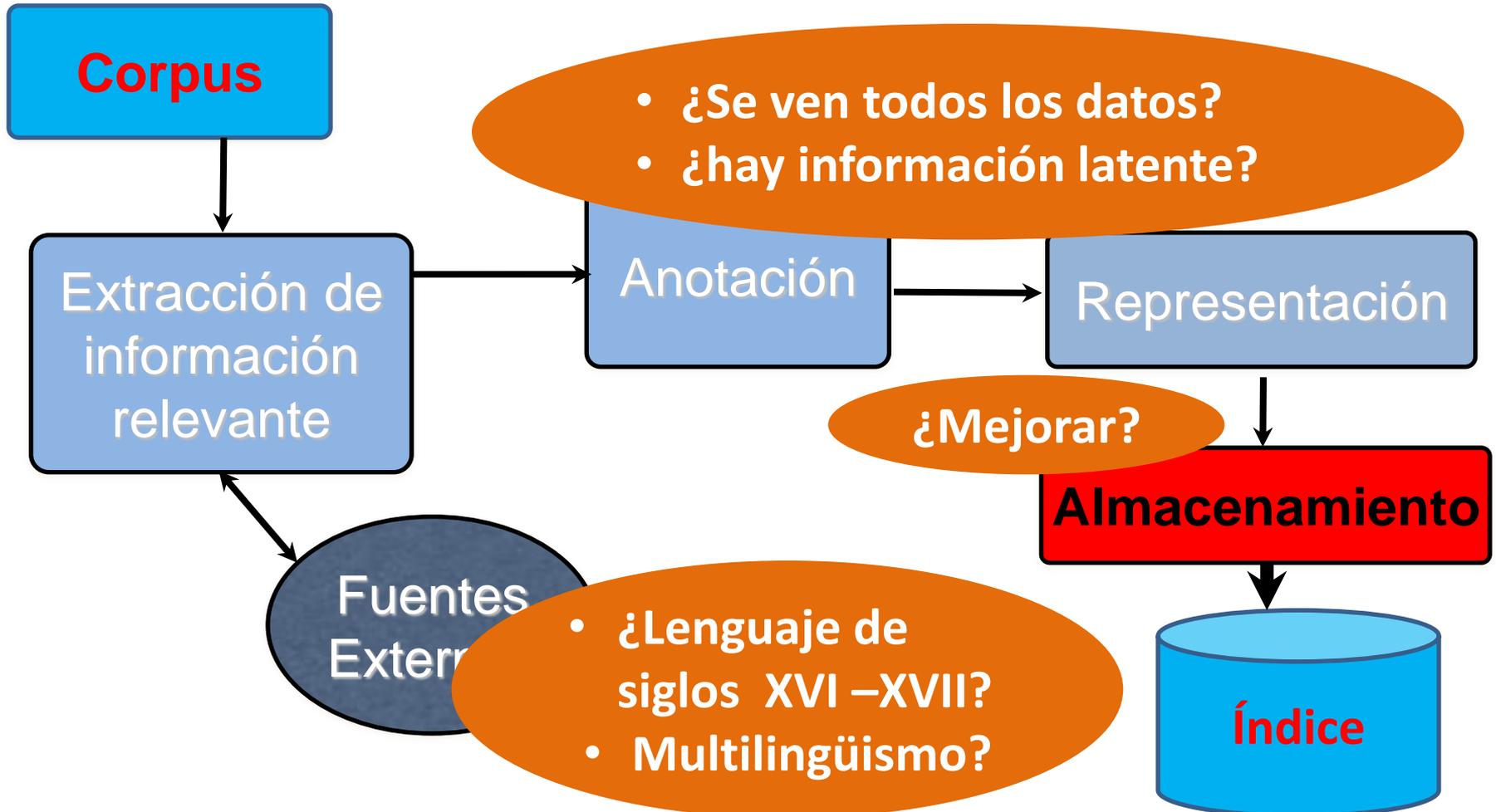
[Pintura de dos medios cañones] [Material gráfico no proyectable]

1587

Idioma: spa

-España -S.XVI AGS. Guerra y Marina, Legajos, 00199, 152. En carta de don Juan de Acuña al Rey, Lisboa, 25 de julio de 1587. Manuscrito sobre papel Tinta negra. Con explicación Referencias: Mapas, planos y dibujos (Años 1503-1805). Volumen I : p. 259 [Pintura de dos medios cañones] [Material gráfico no proyectable] España Juan_de_Acuña Lisboa Lisboa Fotos_Referencias España Lisboa Lisboa Fotos_Referencias Juan_de_Acuña

Hacia el enlazado de datos...



Lattice-Based conceptualization (DIMH project)

Ana García-Serrano
ETSI Infomática, UNED

(In collaboration with Ángel Castellanos and Juan Cigarrán)