

Projecte fi de carrera

Mapa interactiu automàtic de conceptes: Combinació de big data i FCA distribuït

Consultor: David Isern Alarcón

diserna@uoc.edu

Estudiant: Miquel Angel Garcia i Roig

mgarciaroig@uoc.edu

Decembre 2015



Continguts:

- Introducció
- Descripció
- Objectius
- Implementació
- Demostració
- Conclusions



Introducció

- Durant els últims anys, diverses aplicacions sobre l'adquisició de coneixement automàtic o semiautomàtic estan cobrant gran rellevància
- Tant des de fonts estructurades com (cada vegada mes) des de fonts desestructurades
- Abandonant els cercles més acadèmics i incorporant-se paulatinament a la realitat de moltes empreses i institucions
- Tècniques que ja es fan servir hui en dia en àmbits com sistemes de recomanació de compra en aplicacions d'e-commerce o sistemes de digitalització i catàleg de grans museus
- Habitualment les dades sobre les quals extraure el coneixement són de gran volum (i cada vegada esta tendència és més gran)



Descripció (I)

- Ens centrarem en extraure coneixement sobre les relacions (a priori desconegudes) dels objectes sobre un domini donat de una forma automàtica
- Utilitzarem **Formal Concept Analysis (FCA)** com a tècnica: Derivar una jerarquia de conceptes en una ontologia formal des d'una col·lecció d'objectes i les seves propietats
- Cada concepte en la jerarquia representa un conjunt d'objectes compartint els valors “similars” per a un cert conjunt de propietats i cada subconcepte en la jerarquia conté un subconjunt dels objectes per sobre d'ell
- Permiteix “navegar” per objectes “similars” de forma intuïtiva
- Al conjunt d'objectes, juntament amb les propietats d'estudi, així com la funció binària I de pertinença, que indica si per a un objecte aplica una determinada propietat, és anomenada **Formal context**



Descripció (II)

Formal context (representació gràfica)

Objetos\Atributos ²	Compuesto (c)	Par (e)	Impar (o)	Primo (p)	Cuadrado (s)
1	No	No	Sí	No	Sí
2	No	Sí	No	Sí	No
3	No	No	Sí	Sí	No
4	Sí	Sí	No	No	Sí
5	No	No	Sí	Sí	No
6	Sí	Sí	No	No	No
7	No	No	Sí	Sí	No
8	Sí	Sí	No	No	No
9	Sí	No	Sí	No	Sí
10	Sí	Sí	No	No	No



Descripció (III)

- **Formal concept:** conjunt d'objectes (extent) i atributs (intent) que s'extrauen del formal context
- Tots els objectes de l'extent manifesten almenys tots el atributs de l'intent
- Tots el atributs de l'intent son compartits per tots el objectes de l'extent
- En l'exemple anterior:

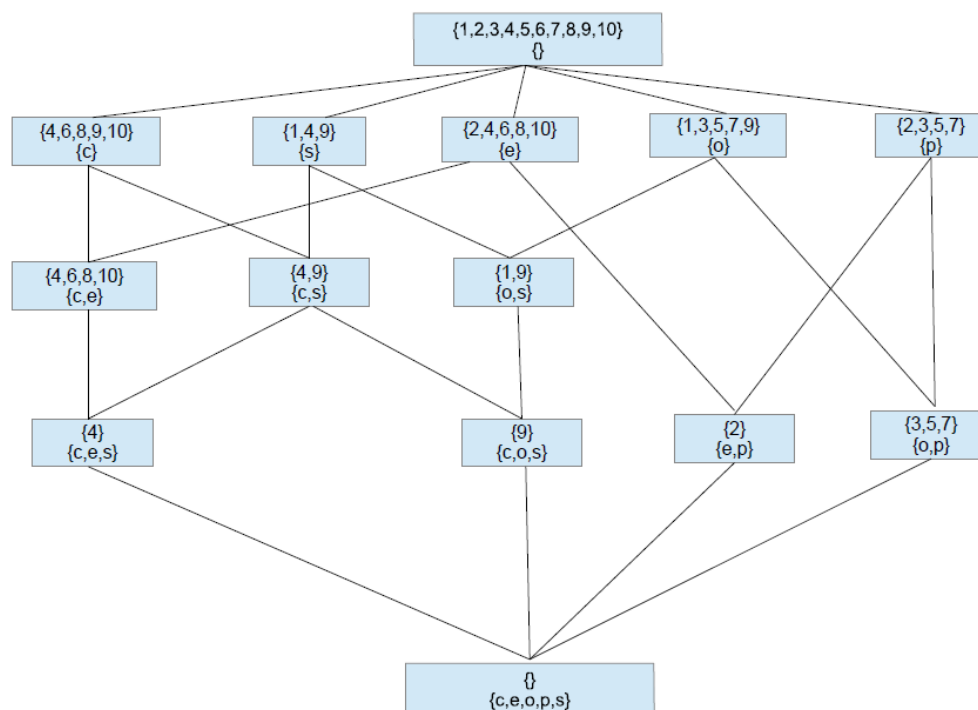
({4, 9}, {c, s})

Tan 4 com 9 es poden descompondre com a sumatori de altres nombres del conjunt, i tots dos també son arrels quadrades; per lo tant comparteixen aquests dos atributs. Cap altres nombres del conjunt manifesten ahora aquests atributs



Descripció (IV)

- **Formal concept lattice:** Graf ordenat dels formal concepts tal que un concepte “hereta” dels conceptes que son super-cojunts del seu intent (objectes) i sub-cojunts del seu extent (atributs). El node arrel es anomenat **màxim**, i el node inferior **mínim**



Descripció (V)

Usos típics de FCA:

- **Ontology alignment.** Minería de dades orientada a la correspondència entre conceptes.
- **Ontology matching.** Troba correspondències entre objectes.
- **Similarity reasoning.** Identificació de conceptes diferents sintàcticament que són semànticament propers (obtenció i integració d'informació a la web semàntica).

En el context d'aquest projecte final de carrera, farem servir una aproximació propera al **ontology matching**, molt utilitzada en processos de digitalització de catàlegs d'alguns museus (Virtual Museum of The Pacific)

Exemple d'utilitat: Aplicació de lloguer de vivendes



Descripció (VI)

Treball previ per tal d'aplicar FCA: Construir el model inicial, o **formal context**.

Transformar el camps de cada objecte, en una serie d'atributs booleans, on '1' o 'true' indica que l'objecte manifesta l'atribut.

Dues aproximacions per tal de generar els atributs:

- **Camps enumeratius.** Generar **un atribut per** cada posible **valor** del domini
- **Camps booleans.** Generar un atribut pel valor cert i un altre pel valor fals
- **Camps numèrics** (o susceptibles de ser transformats, com dates). Utilitzar **algoritmes d'agrupament** sobre els domini de valors del camp.

Hem seleccionat l'algoritme Kmeans con algoritme d'agrupament:

- Mètode no supervisat d'us generalitzat
- Abundant documentació i suport en nombroses eines
- Disponible implementació en Hadoop (Apache Mahout)



Descripció (VII)

Exemple d'estratègia per construir el formal context.

Objecte {

Location.city: València

Information.date:

Information.total_staff: 830

TechData.cooling_forced: false

}



```
{  
  Location.city.ENUM-000: true  
  Location.city.ENUM-001: false  
  Location.city.ENUM-002: false  
  
  Information.date-CLUSTER-001: false  
  Information.date-CLUSTER-002: false  
  Information.date-CLUSTER-003: false  
  Information.date-CLUSTER-UND: true  
  
  Information.total_staff-CLUSTER-001: true  
  Information.total_staff-CLUSTER-002: false  
  Information.total_staff-CLUSTER-003: false  
  Information.total_staff-CLUSTER-UND: false  
  
  TechData.cooling_forced-TRUE: true  
  TechData.cooling_forced-FALSE: false  
  TechData.cooling_forced-UND: false  
}
```



Descripció (VIII)

Domini de problema utilitzat i joc de dades: Dades de reactors nuclears de la IAEA (International Atomic Energy Agency) arreu del mon.

En format excel (entre altres), lliurement accessibles des del lloc web:

<http://nucleus.iaea.org/RRDB/RR/ReactorSearch.aspx?rf=1>



Objectius (I)

El càlcul de FCA no és especialment complicat, però sí costós en temps d'execució (pot arribar a tenir cost exponencial).

Aconseguir realitzar-lo en paral·lel i de forma distribuïda és d'especial rellevància en nom de potencials aplicacions en entorns big data.

Llavors, els nostres objectius són:

- 1) Aprofundir en l'aprenentatge de l'algoritme FCA, tant a nivell teòric com a nivell tècnic.
- 2) Aconseguir una implementació eficient de l'algoritme en termes de temps de computació capaç de ser executada sobre volums mitjans-alts de dades.
- 3) Aplicar de forma adequada tècniques de preparació de dades, transformació de valors i discretització (K-Means en el nostre cas) a la capa ETL en un problema concret.
- 4) Aplicar totes aquesta sèrie de tècniques d'Intel·ligència Artificial per resoldre problemes reals potencials als que s'enfronten empreses i institucions actualment.



Objectius (II)

El nostre enfocament de FCA serà molt proper al *ontology matching*: trobar correspondències entre objectes.

Per plasmar això, desenvoluparem un client web que, donat les dades d'un reactor en concret, mostrarà la informació adient i **enllaços a reactors similars**. Els reactors amb un grau de similitud mes alt, es marcaran d'una forma mes ressaltada a la interfície d'usuari.

Primer seran necessàries:

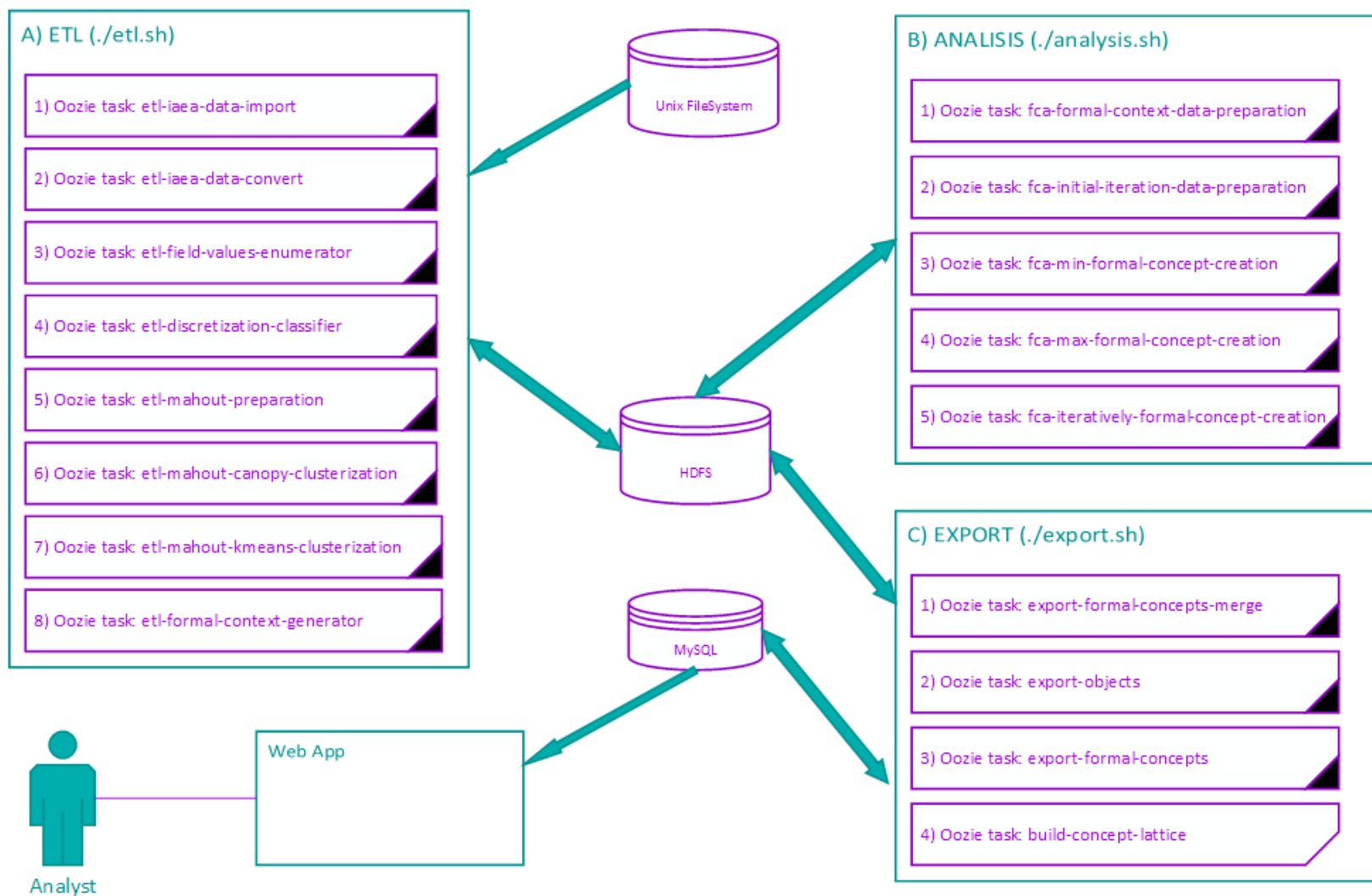
- Una fase d'ETL per generar el formal context inicial
- Una d'anàlisi per generar el concept lattice.
- Finalment una altra per exportar els resultats a una base de dades relacional



Implementació (I)



Implementació (II)



Implementació (III)

Fase ETL

Nom	etl-iaea-data-import
Descripció	Normalitzar format Excel de dades de reactors de la IAEA
Entrada	<code>/user/cloudera/etl/import/ReactorResearch.xls</code>
Eixida	<code>/user/cloudera/etl/import/Merged-ReactorResearch.xls</code>
Tipus	<i>Java action</i>

Nom	etl-iaea-data-convert
Descripció	Generar dades de reactors en format de fitxer de seqüència HDFS
Entrada	<code>/user/cloudera/etl/import/Merged-ReactorResearch.xls</code>
Eixida	<code>/user/cloudera/etl/converted/etl-input.seq</code>
Tipus	<i>Java action</i>



Implementació (IV)

Fase ETL

Nom	etl-field-values-enumerator
Descripció	Generar fitxers HDFS amb mapatges literal - codi numèric per a camps de tipus enumerat
Entrada	<code>/user/cloudera/etl/converted/etl-input.seq</code>
Eixida	<code>/user/cloudera/etl/enumerations</code>
Tipus	<i>Map-Reduce</i>

Nom	etl-discretization-classifier
Descripció	Generar fitxers HDFS amb dades discretitzades per camps de tipus enumerat
Entrada	<code>/user/cloudera/etl/converted/etl-input.seq, /user/cloudera/etl/enumerations/*</code>
Eixida	<code>/user/cloudera/etl/discretized</code>
Tipus	<i>Map-Reduce (map only)</i>

Implementació (V)

Fase ETL

Nom	etl-mahout-preparation
Descripció	Generar fitxers d'entrada per mahout (clusterització K-Means)
Entrada	<code>/user/cloudera/etl/discretized</code>
Eixida	<code>/user/cloudera/etl/mahout-input</code>
Tipus	<i>Map-Reduce</i>

Nom	etl-mahout-canopy-clusterization
Descripció	Generar clusterització inicial de mahout (clusterització canopy)
Entrada	<code>/user/cloudera/etl/mahout-input</code>
Eixida	<code>/user/cloudera/etl/mahout-output/canopy</code>
Tipus	<i>Java action + Map-Reduce (Mahout)</i>

Implementació (VI)

Fase ETL

Nom	etl-mahout-kmeans-clusterization
Descripció	Realitzar la clusterització final <i>K-Means</i> de <i>Mahout</i>
Entrada	<code>/user/cloudera/etl/mahout-input, /user/cloudera/etl/mahout-output/canopy</code>
Eixida	<code>/user/cloudera/etl/mahout-output/kmeans</code>
Tipus	<i>Java action + Map-Reduce (Mahout)</i>
Nom	etl-formal-context-generator
Descripció	Generació del formal context final (fitxer HDFS, atribut => valor booleà)
Entrada	<code>/user/cloudera/etl/discretized, /user/cloudera/etl/mahout-output/kmeans/*-kmeans/clusteredPoints</code>
Eixida	<code>/user/cloudera/etl/formal_context</code>
Tipus	<i>Map-Reduce (map only)</i>

Implementació (VII)

Fase Anàlisi

Función Map

Input: Par $\langle key, value \rangle$ donde key es $\langle B_{0,y} \rangle$ y $value$ es $\langle A, B \rangle$

Result = 0

Para $j = y$ **hasta** $|Y|$ **hacer**

Si $j \in B$ **entonces** continuar;

$C = A \cap \{j\} \downarrow$

$D = C \uparrow$

$result = result \cup \{ \langle B, j \rangle, \langle C, D \rangle \}$

Fin Para

Return result



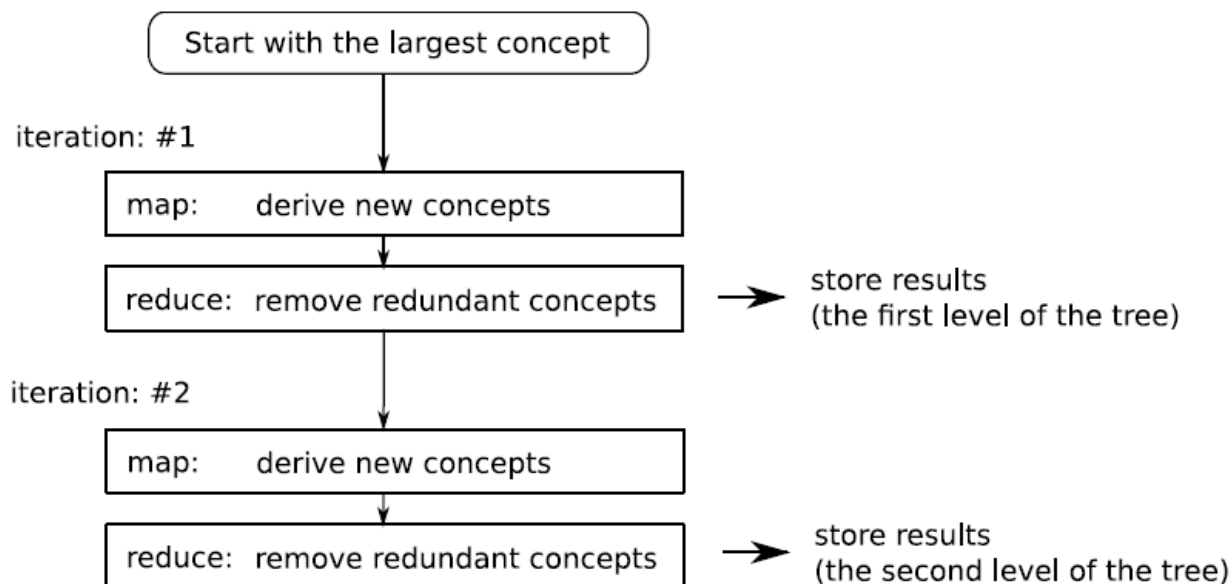
Implementació (VIII)

Fase Anàlisi

Función reduce

$$\text{ReducirConceptos}(\{\langle\langle B, j \rangle, \langle C, D \rangle\rangle\}) = \begin{cases} \{\langle\langle B, j+1 \rangle, \langle C, D \rangle\rangle\}, & \text{si } B \cap Y_j = D \cap Y_j \\ \text{vacío}, & \text{en otro caso} \end{cases}$$

Dónde $Y_j \subseteq Y$ está definido como: $Y_j = \{y \in Y \mid y < j\}$



Implementació (IX)

Fase Anàlisi

Nom	fca-formal-context-data-preparation
Descripció	Generar fitxers amb dades d'objectes compartits per atribut pregenerats
Entrada	<code>/user/cloudera/etl/formal_context</code>
Eixida	<code>/user/cloudera/analysis/data_preparation</code>
Tipus	<i>Map-reduce</i>

Nomb	fca-initial-iteration-data-preparation
Descripció	Generar dades inicials, previs a la primera iteració
Entrada	<code>/user/cloudera/etl/formal_context, /user/cloudera/analysis/data_preparation</code>
Eixida	<code>/user/cloudera/analysis/fca/max_concept</code>
Tipus	<i>Java action</i>



Implementació (X)

Fase Anàlisi

Nomb	fca-min-formal-concept-creation
Descripció	Generar dades de cas especial de node mínim
Entrada	<code>/user/cloudera/etl/formal_context, /user/cloudera/analysis/data_preparation</code>
Eixida	<code>/user/cloudera/analysis/fca/min_concept</code>
Tipus	<i>Java action</i>

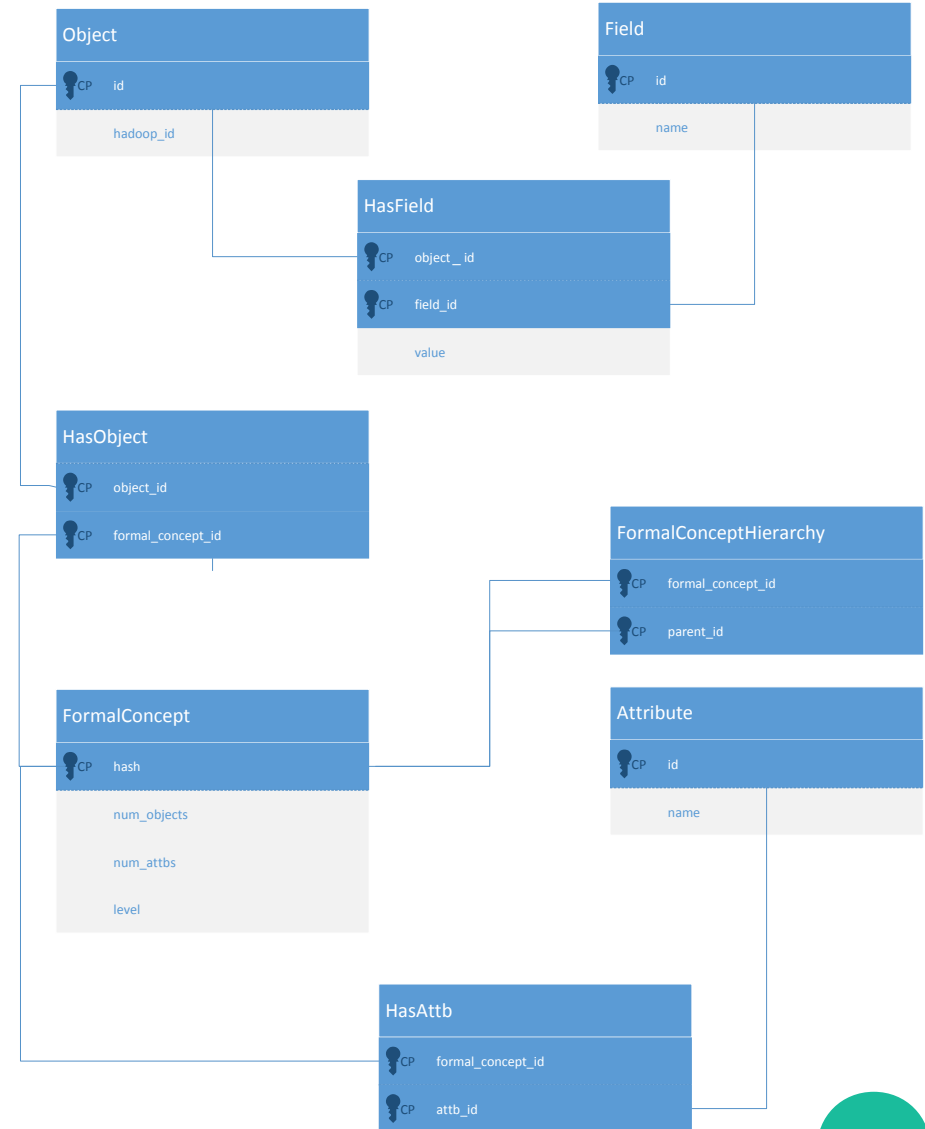
Nom	fca-iteratively-formal-concept-creation
Descripció	Generar <i>formal concepts</i> de forma iterativa
Entrada	<code>/user/cloudera/etl/formal_context, /user/cloudera/analysis/fca/max_concept</code> <code>/user/cloudera/analysis/data_preparation,</code>
Eixida	<code>/user/cloudera/analysis/fca/iterations</code>
Tipus	<i>Java action, que llança jobs Map-reduce fins aconseguir condició de parada</i>



Implementació (XI)

Fase Aplicació d'interfície d'usuari

Model de dades de persistència



Implementació (XII)

Fase Aplicació d'interfície d'usuari

Algoritme per construir el *concept lattice* a la base de dades

Construir *concept lattice*

```
leafConcepts = {maximumFormalConcept()} # root node
level = 0
for conceptAttbSize : 0 .. numberOfAttbs do
  levelInheritance = map() # tipo de datos modo 'diccionario' o array asociativo
  pendingConcepts = findPendingConceptsBySize(conceptAttbSize)
  for concept : pendingConcepts do
    parents = searchForParents(concept, leafNodes)
    levelInheritance[concept] = parents
    insertInheritance(concept,parents)
    markConceptAsProcessed(concept,level)
  end
  if pendingConcepts <> {} then
    updateLeafConcepts(leafConcepts,levelInheritance)
    level = level + 1
  end
end
```



Implementació (XIII)

Fase Aplicació d'interfície d'usuari

Algoritme per construir el *concept lattice* a la base de dades

Función searchForParents (concept, leafConcepts)

```
parents = filterConceptsICanInheritFrom(concept, leafConcepts)
```

```
if parents = {} then
```

```
    parents = closestAncestorsFor(concept, leafConcepts)
```

```
end
```

```
return parents
```



Implementació (XIV)

Fase Aplicació d'interfície d'usuari

Algoritme per construir el *concept lattice* a la base de dades

Función closestAncestorsFor(concept, childrenToCheck)

```
if childrenToCheck = {} then return {}  
parents = findParents(childrenToCheck)  
validAncestors = filterConceptsICanInheritFrom(concept,parents)  
if validAncestors = {} then  
    return closestAncestorsFor(concept,parents)  
end
```



Implementació (XV)

Fase Aplicació d'interfície d'usuari

Algoritme per construir el *concept lattice* a la base de dades

Función updateLeafConcepts (leafConcepts, levelInheritance)

```
for child, parents : levelInheritance do  
  leafConcepts = leafConcepts + {child}  
  for parent : parents do  
    leafConcepts = leafConcepts - {parent}  
  end  
end
```



Implementació (XV)

Fase Aplicació d'interfície d'usuari

Estratègia d'explotació de dades.

Al concepte lattice, els conceptes que hereten van incrementant paulatinament el nombre d'atributs.

D'altra banda, donat un concepte; els conceptes dels que hereta en la jerarquia inclouen tots els objectes que formen part del mateix.

Donç, donat un objecte mostrat a la interfície, haurem de:

- 1) Buscar el concepte que el contingui a la base de dades amb major nombre d'atributs. La resta d'objectes presents al concepte seran els més similars.
- 2) Recorrent la jerarquia d'herència fins a arribar al concepte arrel o màxim ens permetrà obtenir la resta d'objectes amb un grau de similitud menor de manera progressiva.



Implementació (XVI)

Fase Aplicació d'interfície d'usuari

Pàgina principal

Mapa interactivo automático de conceptos: Combinación de Big Data y FCA distribuido

Listado de objetos

[05155f34-05a0-48ae-969e-7d8e6455d2a7](#)
[05742284-b638-413a-992e-af7149b54be4](#)
[078c7fb1-650a-4897-be21-68b6bef9542d](#)
[08788fd0-e062-4032-afe2-f9f0e257f6a5](#)
[087a403f-b071-4c1f-9316-1ec607378c81](#)
[0936ac21-7bca-449b-b097-01c27b7498e4](#)
[0a249935-cad6-4a2b-a3da-3f9411e0af06](#)
[0bcaaafe-ac28-4a50-b24e-ceb519f81917](#)
[0f9ff1ed-bdc4-4d0b-8660-667f6205b2d3](#)
[16a4b147-c8fe-4c52-9e9e-c53da83134bd](#)
[177274d8-8496-4cbb-8ede-662de7d37102](#)

Por favor, seleccione un objeto desde el panel de la izquierda para comenzar el análisis



Implementació (XVIII)

Fase Aplicació d'interfície d'usuari

Vista d'objecte seleccionat

Mapa interactivo automático de conceptos: Combinación de Big Data y FCA distribuido

Listado de objetos

[05155f34-05a0-48ae-969e-7d8e6455d2a7](#)
[05742284-b638-413a-992e-af7149b54be4](#)
[078c7fb1-650a-4897-be21-68b6bef9542d](#)
[08788fd0-e062-4032-afe2-f9f0e257f6a5](#)
[087a403f-b071-4c1f-9316-1ec607378c81](#)
[0936ac21-7bca-449b-b097-01c27b7498e4](#)
[0a249935-cad6-4a2b-a3da-3f9411e0af06](#)
[0bcaafe-ac28-4a50-b24e-ceb519f81917](#)
[0f9ff1ed-bdc4-4d0b-8660-667f6205b2d3](#)
[16a4b147-c8fe-4c52-9e9e-c53da83134bd](#)
[177274d8-8496-4cbb-8ede-662de7d37102](#)
[19382775-f900-447c-a2d6-fefc0779b8af](#)
[1a39a89f-b331-4ba5-a6be-c50d89e25eeb](#)
[1e15d776-2b87-4542-b3ce-a5251f571d10](#)
[22f388e7-36d8-4853-83ed-5035052557ce](#)
[23f60945-3c0c-47b9-b8e6-c425c8e19ef7](#)
[2411a0b4-aa68-424e-912d-fd57d2f3396e](#)
[243eebba-4db4-4930-8cf7-95ba3d23b233](#)
[2495abe4-dc45-465a-b444-8dc9b30a79d2](#)
[26ef3a5d-e94b-4be7-9fe1-dff7f1d40d10](#)
[27bbbe84-52be-489b-9a62-d5bb89c6a4f0](#)
[2831de35-1aee-4a22-9fa9-118b69692ada](#)
[285a3fa3-7c87-4a35-84d3-8bc829386621](#)

Objeto: 05155f34-05a0-48ae-969e-7d8e6455d2a7

▼ Facility

IAEA_code	es0001
name	jen-1 mod

- ▶ Category
- ▶ Status
- ▶ Location
- ▶ Information
- ▶ TechData
- ▶ Utilization

Objetos similares: De mayor a menor similitud

- ▶ Coincidencias de 30 atributos
- ▶ Coincidencias de 29 atributos
- ▶ Coincidencias de 28 atributos
- ▶ Coincidencias de 27 atributos
- ▶ Coincidencias de 26 atributos
- ▶ Coincidencias de 25 atributos
- ▶ Coincidencias de 24 atributos
- ▶ Coincidencias de 23 atributos
- ▶ Coincidencias de 16 atributos
- ▶ Coincidencias de 15 atributos

Implementació (XIX)

Fase Aplicació d'interfície d'usuari

Vista de coincidències

Mapa interactivo automático de conceptos: Combinación de Big Data y FCA distribuido

Listado de objetos

[05155f34-05a0-48ae-969e-7d8e6455d2a7](#)
[05742284-b638-413a-992e-af7149b54be4](#)
[078c7fb1-650a-4897-be21-68b6bef9542d](#)
[08788fd0-e062-4032-afe2-f9f0e257f6a5](#)
[087a403f-b071-4c1f-9316-1ec607378c81](#)
[0936ac21-7bca-449b-b097-01c27b7498e4](#)
[0a249935-cad6-4a2b-a3da-3f9411e0af06](#)
[0bcaafe-ac28-4a50-b24e-ceb519f81917](#)
[0f9ff1ed-bdc4-4d0b-8660-667f6205b2d3](#)
[16a4b147-c8fe-4c52-9e9e-c53da83134bd](#)
[177274d8-8496-4cbb-8ede-662de7d37102](#)
[19382775-f900-447c-a2d6-fefc0779b8af](#)
[1a39a89f-b331-4ba5-a6be-c50d89e25eeb](#)
[1e15d776-2b87-4542-b3ce-a5251f571d10](#)
[22f388e7-36d8-4853-83ed-5035052557ce](#)
[23f60945-3c0c-47b9-b8e6-c425c8e19ef7](#)
[2411a0b4-aa68-424e-912d-fd57d2f3396e](#)
[243eebba-4db4-4930-8cf7-95ba3d23b233](#)
[2495abe4-dc45-465a-b444-8dc9b30a79d2](#)
[26ef3a5d-e94b-4be7-9fe1-dff7f1d40d10](#)
[27bbbe84-52be-489b-9a62-d5bb89c6a4f0](#)
[2831de35-1aee-4a22-9fa9-118b69692ada](#)
[285e3fe3-7c87-4a35-8dd3-8bc829386621](#)
[29d90dd9-0bf5-4dcc-9402-079f43187348](#)
[2a6b2bc0-4df6-4f00-bdca-61c12c15015a](#)
[2a759fd6-0eec-4333-b9c3-2077418af813](#)

Objeto: 05155f34-05a0-48ae-969e-7d8e6455d2a7

▼ Facility
 IAEA_code es0001
 name jen-1 mod
 ▶ Category
 ▶ Status
 ▶ Location
 ▶ Information
 ▶ TechData
 ▶ Utilization

Objetos similares: De mayor a menor similitud

▼ Coincidencias de 30 atributos

▼ Objetos similares: 17

Objetos

[05155f34-05a0-48ae-969e-7d8e6455d2a7](#) 285e3fe3-7c87-4a35-8dd3-8bc829386621 365b1969-41d1-4bf8-9de5-378c26379295 3858a08b-b9f8-43d0-ad0b-57d88f949e79 3fd1332-4388-4879-a212-ed833fef5cbdb 4c56c4b9-bda4-4acf-ad84-0877d5f22706 50366883-7a26-4d3d-bf67-39bc7813ff65 5c371f53-61c2-46ed-98c7-41cf9ae0cdee 738d9e61-d338-41e8-8f8b-49edc1dbb90b 8687b715-8555-49be-b938-e8d7069eca2c 8db07edb-cba2-466f-99b4-4ea7d4132847 90f43861-d41a-41dc-a2d9-e707d933733e 9ead7e1f-cee7-472c-9b87-979869a7de18 afa38522-dfcc-4822-9db4-071df8028a89 b155231d-2854-485e-8290-15d3227e2854 c5dd9780-91a6-4b0d-bf89-6e8c4e0e4e27 e3f3c30a-7108-4d68-bedd-0fc4f8e54ef7

Campos coincidentes

Category.category Experimental.irradiation_channels Experimental.reflector_max_flux TechData.coolant_material TechData.max_flux_pulsed_fast TechData.max_flux_pulsed_thermal Utilization.days_per_week Utilization.experimenters_number Utilization.geochronology_methods Utilization.geochronology_samples_per_year Utilization.hours_per_day Utilization.isotope_total_activity_per_year Utilization.isotopes Utilization.materials Utilization.mw_days_per_year Utilization.neutron_activation_analysis_methods Utilization.neutron_activation_analysis_samples_per_year Utilization.neutron_capture_therapy Utilization.neutron_capture_therapy_patient_per_year Utilization.neutron_radiography

Implementació (XX)

Fase Aplicació d'interfície d'usuari

Nombre	export-formal-concepts-merge
Descripció	Unificar tots els fitxers de formal concepts generats
Entrada	<code>/user/cloudera/analysis/fca/**/*</code>
Eixida	<code>/user/cloudera/export/mergedFormalConcepts</code>
Tipus	<i>Java action</i>

Nombre	export-objects
Descripció	Exportar els objectes (reactors) a la base de dades
Entrada	<code>/user/cloudera/etl/converted/etl-input.seq, /user/cloudera/etl/discretized</code>
Eixida	<code>Tablas object,field,has_field</code>
Tipus	<i>Java action</i>



Implementació (XXI)

Fase Aplicació d'interfície d'usuari

Nom	export-formal-concepts
Descripció	Exportar els formal concept a la base de dades
Entrada	<code>/user/cloudera/export/mergedFormalConcepts,</code> Taules <code>object,field,has_field</code>
Eixida	Taules <code>formal_concept, attribute, has_object, has_attb, formal_concept_hierarchy</code>
Tipus	<i>Java action</i>

Nom	build-concept-lattice
Descripció	Construir jerarquia "concept lattice" entre els conceptes exportats a la base de dades
Entrada	Taules <code>object, formal_concept, attribute, has_object, has_attb</code>
Eixida	Taules <code>formal_concept, formal_concept_hierarchy</code>
Tipus	<i>Java action</i>



Demostració

Pasarem ara a mostrar la execució, tant al cluster hadoop com el client web final



Conclusions (I)

Desenvolupant aquest projecte, hem tingut la oportunitat de:

- Aprofundir sobre diversos paradigmes de la IA, concretament sobre establiment de coneixement a través de les relacions entre objectes
- Concretament hem aprofundit en la tècnica del **formal concept analysis**, que consisteix en derivar una jerarquia de conceptes en una ontologia formal des d'una col·lecció d'objectes i les seves propietats; on la estructura jeràrquica de l'ontologia ens permet extreure coneixement sobre la interrelació dels elements que la formen amb facilitat.
- Per això hem assimilat conceptes com el **formal context**, ; així com el conjunt d'**atributs** d'estudi i una *funció de pertinença*. A partir d'aquesta estructura hem après a identificar els **formal concepts**, o subconjunts d'elements derivats a partir del model inicial de l'formal context que els fan ideals per a raonar sobre similitud; així com la seva jerarquitització i ordenació en el graf final denominat **concept lattice**



Conclusions (II)

- Així mateix, hem detallat els principals usos de l'FCA: **ontology alignment**, **ontology matching** i **Similarity reasoning**; i hem enquadrat el nostre problema en concret dins el ontology matching.
- Hem estudiat el **cost** de les implementacions més habituals de la FCA, veient com pot ser **exponencial**; pel que hem posat en valor la possibilitat d'una implementació sobre un algoritme adaptat a un framework de **processament distribuït** com és **Hadoop**
- Centrats en aquest entorn de computació, hem aprofitat implementacions existents de diversos algoritmes de IA -concretament **clusterització Kmeans**- amb **Apache mahout**, pensades per operar sobre grans volums de dades d'entrada
- Hem tingut l'oportunitat de comprovar la complexitat de desenvolupament en un entorn d'aquesta naturalesa, i de l'alt volum de dades que es poden generar.



Conclusions (III)

- També hem constatat com la generació del concepte lattice no és trivial quan es treballa amb un volum de dades mitjà / gran; i hem adaptat un algoritme propi susceptible de ser aplicat per tractar formalment conceptes carregats des de bases de dades relacionals.
- Finalment, hem après a conceptualitzar arquitectures complexes que combinin diferents aproximacions per tractar problemes on la resolució se sustenta sobre algorismes i tecnologies heterogènies.



Conclusions (IV)

Futures millores i ampliacions

- Extraure algunes configuracions del mòdulo del backend, per tal de que la solució adoptada se 100% independent del domini del problema
- Optimitzar el consum de memòria en algunes accions al cluster Hadoop
- Extendre l'algoritme map-reduce de generació de conceptes per incloure també la herència i ordenació de conceptes
- Adaptar l'algoritme a la versió difusa de FCA (Fuzzy-FCA)

