Visualización de Esquemas y Datos en Bases de Datos NoSQL

Trabajo Fin de Master

Alberto Hernández Chillón

Tutores

Jesús García Molina Diego Sevilla Ruiz

Facultad de Informática Universidad de Murcia

Julio 2016

Índice

- 1 Introducción y análisis de objetivos
- 2 Proceso de inferencia previo
- Visualización de esquemas
- Visualización de datos
- Conclusiones y vías futuras

Índice

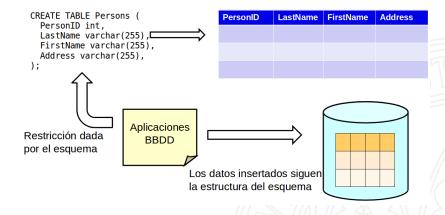
- 1 Introducción y análisis de objetivos
- Proceso de inferencia previo
- 3 Visualización de esquemas
- 4 Visualización de datos
- Conclusiones y vías futuras

Introducción y contexto

Introducción al problema

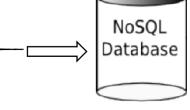
- Creciente número de aplicaciones modernas que exigen unas prestaciones y funcionalidad particulares:
 - Manejo de gran volumen de información
 - Almacenamiento de datos complejos
 - Procesamiento escalable
- Los sistemas relacionales tradicionales presentan limitaciones para satisfacer estas condiciones
- Adopción de sistemas no relacionales NoSQL para paliar la situación

Bases de datos NoSQL (I)



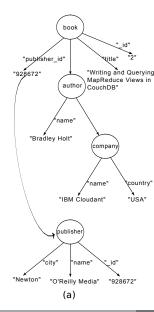
Bases de datos NoSQL (II)

```
'person id': '123',
'type': 'Person',
'lastName': 'Rush',
'firstName': 'Christopher',
'address': 'C/ Gran Vía, 13, Madrid'
'person id': '456',
'type': 'Person',
'lastName': 'England',
'firstName': 'Wayne',
'address': {
    'street': 'Av. Pinos, 24',
    'city': Murcia
'person id': '789',
'type': 'Person',
'lastName': 'Sevilla',
'firstName': 'Diego',
'address': 'Ronda Norte, 15, Murcia',
'age': 35
```



- Información no uniforme
- Existencia de distintas versiones

Bases de datos NoSQL (III)



```
"rows": [
  "content": {
     "chapters": 33
     "pages": 527
  "authors": [
     "company": {
        "country": "USA".
        "name": "IRM"
     "name": "Grady Booch".
     " id": "210"
     "company": {
        "country": "USA".
        "name": "IBM"
     "name": "James Rumbaugh".
     " id": "310"
     "country": "USA".
     "company": "Ivar Jacobson
                  Consulting (IJC)",
     "name": "Ivar Jacobson".
     " id": "410"
   "type": "book".
    "year": 2013,
    "publisher id": "345679".
    "title": "The Unified Modeling
           Language".
    " id": "1"
  "discipline": "software engineering",
  "issn": [
      "0098-5589".
      "1939-3520"
  "name": "IEEE Trans. on Software
          Engineering".
  "type": "journal",
   id": "11"
                                        1 11
                                   (b)
```

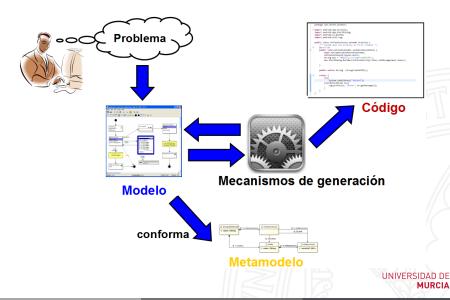
```
"name": "Automated Software Engineering".
"issn": [
  "0928-8910"
  "1573-7535"
"discipline": "software engineering".
"type": "journal",
" id": "12".
"number":10515
"city": "Barcelona"
"name": "Omega".
"type": "publisher",
" id": "123451"
"citv": "Newton".
"name": "O'Reilly Media"
"type": "publisher".
" id": "928672"
"type": "book".
"author": {
   " id": "101".
   "name": "Bradley Holt",
   "company": {
  "country": "USA",
     "name": "IBM Cloudant",
"title": "Writing and Querying MapReduce
       Views in CouchDB".
"publisher id": "928672".
" id": "2"
  "name": "Addison-Wesley",
  "type": "publisher".
  " ld": "345679"
"type": "publisher",
"iournals": [
   "11"
   "12"
"name": "IFFF Publications"
  " id": "907863"
                              UNIVERSIDAD DE
                                          MURCIA
```

Análisis de objetivos

Análisis de objetivos

- Un informe de *DataDiversity* del 2015 señala utilidades que serán esenciales en un futuro próximo: Visualización de modelos, generación de código a partir de esquemas y manejo de metadatos
- El objetivo es el diseño e implementación de herramientas que permitan:
 - Visualizar esquemas NoSQL
 - Teniendo en cuenta las posibles versiones de cada esquema
 - Visualizar datos NoSQL
 - Clasificando estos datos de acuerdo a su entidad y versión

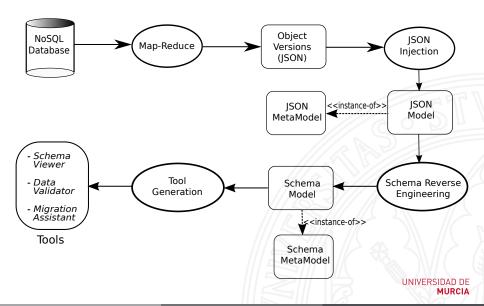
Solución MDE



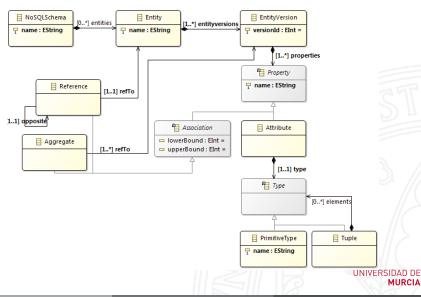
Índice

- 1 Introducción y análisis de objetivos
- 2 Proceso de inferencia previo
- Visualización de esquemas
- 4 Visualización de datos
- Conclusiones y vías futuras

Proceso de inferencia previo (I)



Proceso de inferencia previo (II)

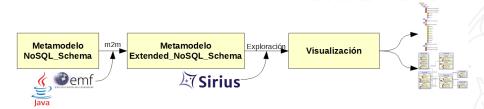


Índice

- 1 Introducción y análisis de objetivos
- Proceso de inferencia previo
- 3 Visualización de esquemas
- 4 Visualización de datos
- 5 Conclusiones y vías futuras

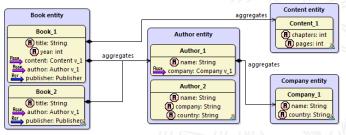
Herramienta desarrollada y proceso seguido

- Interés en visualizar esquemas y versiones de esquema
- Integración con la infraestructura EMF, eficiencia, amigabilidad
- Diseño de distintas vistas de descripción de un modelo
- Navegabilidad entre vistas



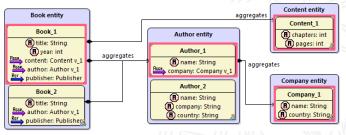
Diseño del metamodelo Extended_NoSQL_Schema

- Existencia de distintas versiones de esquema: Agrupación de entidades y versiones a partir de una versión raíz
- Extensión a NoSQL_Schema: Se define la clase SchemaVersion y sus relaciones



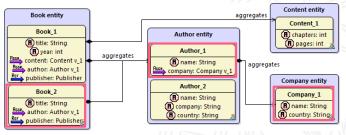
Diseño del metamodelo Extended_NoSQL_Schema

- Existencia de distintas versiones de esquema: Agrupación de entidades y versiones a partir de una versión raíz
- Extensión a NoSQL_Schema: Se define la clase SchemaVersion y sus relaciones



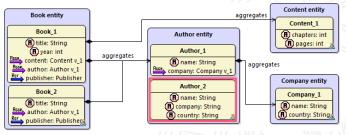
Diseño del metamodelo Extended_NoSQL_Schema

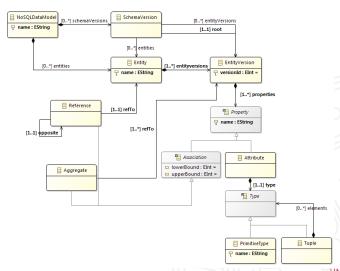
- Existencia de distintas versiones de esquema: Agrupación de entidades y versiones a partir de una versión raíz
- Extensión a NoSQL_Schema: Se define la clase SchemaVersion y sus relaciones

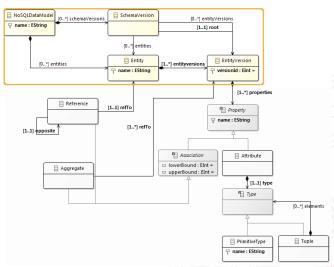


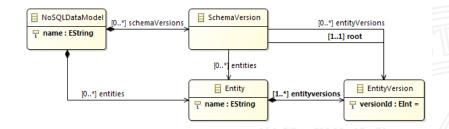
Diseño del metamodelo Extended_NoSQL_Schema

- Existencia de distintas versiones de esquema: Agrupación de entidades y versiones a partir de una versión raíz
- Extensión a NoSQL_Schema: Se define la clase SchemaVersion y sus relaciones



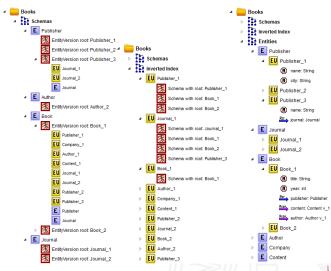


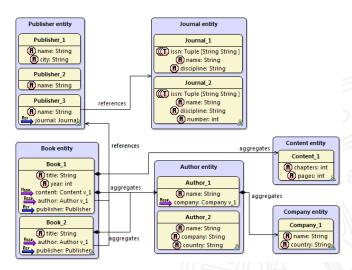


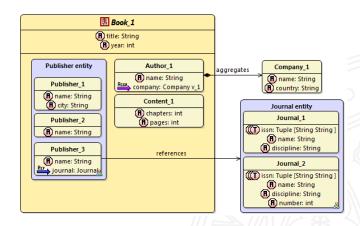


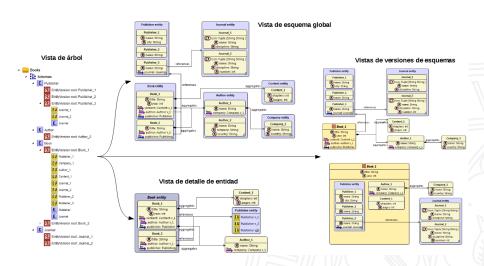
Diseño e implementación de las vistas con Sirius

- Herramienta de creación de DSLs gráficos integrada en Eclipse
- Creación de viewpoints con vistas definidas para un metamodelo
- Asociación entre un elemento del metamodelo y una representación gráfica o relación entre elementos
- Definición de acciones para la creación, edición y navegabilidad entre elementos
- Generación automática de un editor embebido en Eclipse









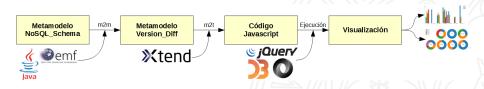
Índice

- 1 Introducción y análisis de objetivos
- 2 Proceso de inferencia previo
- 3 Visualización de esquemas
- Visualización de datos
- Conclusiones y vías futuras

Visualización de datos (I)

Herramienta desarrollada y proceso seguido

- Interés en visualizar documentos JSON extraidos de una base de datos NoSQL
- Clasificación de los datos obtenidos por entidad y versión
- Ordenación y visualización de los datos
- Diseño de distintas vistas de clasificación de datos



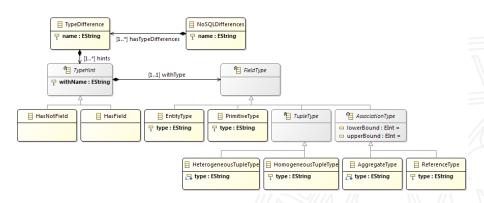
Visualización de datos (II)

Diseño del metamodelo Version Diff

- Necesidad de obtener un método para, dado un objeto de la base de datos, identificar su entidad y versión
- Metamodelo de diferencias: Recoge los campos de cada objeto con clases HasField y HasNotField
 - Un objeto se identifica por sus campos HasField definidos
 - Un objeto se distingue de versión por sus campos HasNotField

```
Entity type: 'Person'
                                                                                                              Entity type: 'Person',
                                   HasField: 'lastName'.
                                                                                                            HasField: 'lastName',
'person id': '123'.
                                                                              'person id': '678',
                                     Primitive type: 'string',
                                                                                                              Primitive type: 'string'
'type': 'Person'.
                                   HasField: 'firstName'.
                                                                                                            HasField: 'firstName',
'lastName': 'Rush',
                                                                              'lastName': 'Nielsen'.
                                    Primitive type: 'string'.
                                                                                                              Primitive type: 'string',
'firstName': 'Christopher'
                                                                             'firstName': 'Terese'.
                                  HasField: 'address',
                                                                                                            HasField: 'age'.
'address': 'C/ Gran Vía.
                                    Primitive type: 'string'
                                                                                                              Primitive type: 'int',
  13. Madrid'.
                                  HasNotField: 'age',
                                                                                                            HasNotField: 'address'.
                                     Primitive type: 'int'
                                                                                                              Primitive type: 'string
```

Visualización de datos (III)



Visualización de datos (IV)

Generación del código de clasificación

- Transformación del modelo de diferencias a código JavaScript
- Por cada versión se genera un método que recibe un objeto como parámetro y comprueba sus campos
 - Campo HasField: Debe existir un campo con un determinado nombre y tipo
 - Campo HasNotField: No debe existir un campo con un determinado nombre y tipo
- Este método devuelve un valor booleano indicando si el objeto dado es de una versión determinada
- Iteración de cada objeto JSON con cada método de clasificación: Se obtiene la entidad y versión del objeto

Visualización de datos (V)

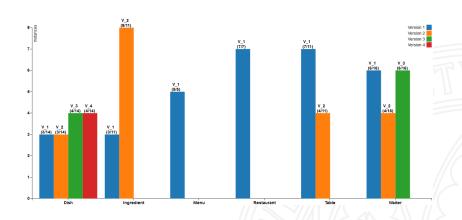
```
var DiffMethodsExclusive =
    isOfExactType Journal 1: function (obj)
      if (!("issn" in obj) || !(obj.issn.constructor === Array)
        (!checkAllOf(obi.issn. "String")))
          return false:
       if (!("name" in obi) || !(typeof obi.name === "string"))
          return false:
       if (!("discipline" in obj) || !(typeof obj.discipline === "string"))
10
          return false;
11
       if (!("type" in obj) || !(typeof obj.type === "string")
12
        || (obj.type !== "Journal"))
13
          return false:
       if (!("owner" in obj) || (typeof obj.owner === "string" && false)
        | | (obj.owner.constructor = Array
16
          && (1 > obj.owner.size || !checkAllOf(obj.owner, "string"))
17
        || (typeof obj.owner !== "string"
          && obj.owner.constructor !== Array)))
19
          return false:
20
       if ("number" in obj && !(!(typeof obj.number === "number")
        || !(obj.number %1 == 0)))
22
          return false;
      return true:
26
    }.
27
```

Visualización de datos (VI)

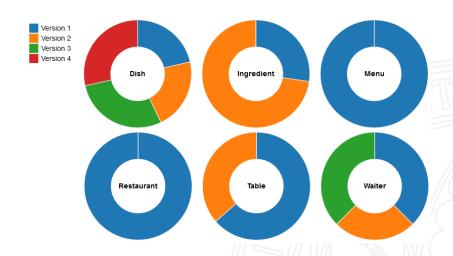
Generación de código de visualización

- Del proceso de clasificación se obtiene una estructura de objetos ordenados
- Visualización de estos objetos con funciones D3.js
- El código de visualización y las funciones auxiliares son independientes del modelo de entrada
- Fácilmente extensible con otros métodos de visualización

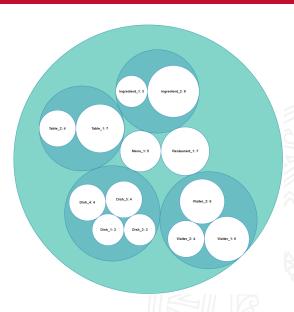
Visualización de datos (VII)



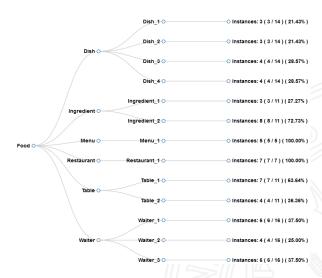
Visualización de datos (VIII)



Visualización de datos (IX)



Visualización de datos (X)



Índice

- 1 Introducción y análisis de objetivos
- Proceso de inferencia previo
- Visualización de esquemas
- 4 Visualización de datos
- Conclusiones y vías futuras

Conclusiones y vías futuras(I)

Conclusiones

- Desarrollada una de las primeras propuestas de visualización NoSQL considerando versiones
- Herramienta para visualizar esquemas y versiones de esquemas de bases de datos con Sirius en Eclipse
 - Vista de árbol y esquema global
 - Vistas de versión de esquema y detalle de entidad
- Herramienta para visualizar datos ordenados por entidad y versión con Javascript y D3.js
 - Diagramas de barras, burbujas, dónuts y árboles
 - Fácilmente extensible
- La utilización de técnicas MDE ha reducido el esfuerzo de implementación de las herramientas

Conclusiones y vías futuras(II)

Vías futuras

- Extensión de la herramienta Sirius para permitir modelado de esquemas de base de datos NoSQL
- Creación de una herramienta independiente de la plataforma
- Visualización de datos como servicio web
- Procesamiento de visualización de datos en dos pasos
- Explotación del metamodelo *Version_Diff* para inferir información de atributos, progresiones y estadísticas

Visualización de Esquemas y Datos en Bases de Datos NoSQL

¡Gracias por la atención!

¿Preguntas?

Alberto Hernández Chillón alberto hernandez 10 um. es