KnowledgeStore
— Scalable Framework for Interlinking Text and Knowledge —

Marco Rospocher
Fondazione Bruno Kessler (FBK)
rospocher@fbk.eu :: @marcorospocher
• A scalable, fault-tolerant, and Semantic Web grounded storage system to jointly store, manage, retrieve, and query, both structured and unstructured data
KnowledgeStore

KnowledgeStore
Scalable storage for text and RDF data

Event extraction

NAF

RDF

Mentions to Instances

IDAP

NLP technologies
Semantic Web technologies
Visualizations
Among a collection of news articles, a user is interested in retrieving all 2014 articles reporting statements of a 20th century US president where he is positively mentioned as “commander-in-chief”.
• Among a collection of news articles, a user is interested in retrieving all 2014 articles reporting statements of a 20th century US president where he is positively mentioned as “commander-in-chief”.
"Stripes," Dole, 1996

He (Bill Clinton) quickly became the first civilian commander-in-chief to salute his marine guards while entering or exiting an aircraft.
KnowledgeStore

Exploitation

- Enhanced applications (e.g., decision support systems)
- Developing, debugging, training, and evaluating NLP and knowledge processing tasks
- Reasoning on Extracted Information (e.g., on Events)
- Text Exploration
Knowledge Store
Architectural View

Client-side

Any application
(HTTP access to the KS, possibly exploiting SPARQL client libraries)

Java applications
(KnowledgeStore Java client)

Server-side

Hadoop HDFS
(name & data nodes)

Hbase / ElasticSearch
(multiple server nodes)

Zookeeper
(mult. nodes)
distributed synchronization

Representation

Resource

Mention

KnowledgeStore Frontend Server

SPARQL endpoint

CRUD endpoint

Virtuoso
(single node)

transaction manager

OMID
(single node)
transaction manager
Looking through the glass box: the User Interface

KnowledgeStore UI

KnowledgeStore/1.7-SNAPSHOT

This is the entry point of the KnowledgeStore Web API.

Please refer to the KnowledgeStore Web site and, in particular, to the Web API reference documentation for information on how to invoke it; an automatically generated WADL descriptor is also available.

A small UI can be used to explore the contents in the KnowledgeStore, either by looking up specific URIs or by evaluating SPARQL queries (refer to upper right menu).

System status: 41d2h21m uptime, 0% gc; 417/1074/816 MB memory used/peak-committed; 44/53/354728 threads active/peak/started
Knowledge Store

Looking through the glass box: the User Interface

Lookup of a resource
Looking through the glass box: the **User Interface**

Lookup of a mention
Looking through the glass box: the **User Interface**

**SPARQL query**

```sparql
WHERE {
  ?event a sem:Event, eso:JoiningAnOrganization .
  ?event rdf:type ?event_label .
  ?event eso:employment-employer dbpedia:Kia_Motors .
  ?event sem:hasTime ?time .
  ?time owl:inDateTime ?time_owl .
  ?time_owl owl:year ?year ; owl:month ?month ; owl:day ?day .
}
ORDER BY ?year ?month ?day.
```
Looking through the glass box: the **User Interface**

**SPARQL query**

```
WHERE {
  ?event a sem:Event, eso:JoiningAnOrganization .
  ?event eso:employment-employer dbpedia:Kia_Motors .
  ?event sem:hasTime ?time .
  ?time owltime:inDateTime ?time_owl .
  ?time_owl owltime:year ?year ;
  owltime:month:month ?month ;
  owltime:day:day ?day .
}
ORDER BY ?year ?month ?day
```

Check out the UI demonstration video: [https://youtu.be/YVOQaljLta4](https://youtu.be/YVOQaljLta4)
KnowledgeStore

Reasoning on Events

• Inferring Knowledge **not Explicitly Mentioned** in Text (powered by ESO)
• Example: “Kia has hired Peter Schreyer as chief design officer.”
Inferring Knowledge **not Explicitly Mentioned** in Text (powered by ESO)

- Example: “Kia has hired Peter Schreyer as chief design officer.”
Inferring Knowledge **not Explicitly Mentioned** in Text (powered by ESO)

Example: “Kia has hired Peter Schreyer as chief design officer.”

**KnowledgeStore**

Reasoning on Events

- Inferring Knowledge **not Explicitly Mentioned** in Text (powered by ESO)
- Example: “Kia has hired Peter Schreyer as chief design officer.”
• Inferring Knowledge **not Explicitly Mentioned** in Text (powered by ESO)
• Example: “Kia has hired Peter Schreyer as chief design officer.”
## Instances Populated During the Project

<table>
<thead>
<tr>
<th>Topic</th>
<th>Period</th>
<th>News Providers</th>
<th>Language</th>
<th>Populated in</th>
<th>Pipeline Version</th>
<th>Mentions</th>
<th>Events</th>
<th>Entities</th>
<th>Persons</th>
<th>in DBpedia</th>
<th>Organizations</th>
<th>in DBpedia</th>
<th>Locations</th>
<th>in DBpedia</th>
<th>Triples</th>
<th>from Mentions</th>
<th>from DBpedia</th>
<th>distilled from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>MEANTIME</td>
<td>WikiNews (Ver. 1)</td>
<td>WikiNews (Ver. 2)</td>
<td>Cars (Ver. 1)</td>
<td>FIFA WorldCup</td>
<td>Dutch Parliament</td>
<td>Cars (Ver. 2)</td>
<td>Cars (Ver. 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>---------------</td>
<td>------------------</td>
<td>---------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td>English</td>
<td>English</td>
<td>English</td>
<td>Dutch</td>
<td>English</td>
<td>English</td>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipeline Version</td>
<td>3.0</td>
<td>2.0</td>
<td>3.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News Articles</td>
<td>120</td>
<td>18,510</td>
<td>19,755</td>
<td>63,635</td>
<td>212,256</td>
<td>597,530</td>
<td>1,259,74</td>
<td>2,316,158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mentions</td>
<td>35,237</td>
<td>2,629,176</td>
<td>5,206,202</td>
<td>9,110,683</td>
<td>76,165,114</td>
<td>9,231,113</td>
<td>205,114,71</td>
<td>842,639,827</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events</td>
<td>3,333</td>
<td>624,439</td>
<td>632,704</td>
<td>1,783,991</td>
<td>9,387,356</td>
<td>5,363,498</td>
<td>25,156,57</td>
<td>42,296,287</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entities</td>
<td>339</td>
<td>45,592</td>
<td>40,314</td>
<td>428,700</td>
<td>858,982</td>
<td>111,579</td>
<td>1,967,15</td>
<td>2,263,156</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons</td>
<td>82</td>
<td>19,677</td>
<td>17,617</td>
<td>199,999</td>
<td>403,021</td>
<td>43,546</td>
<td>729,79</td>
<td>895,541</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in DBpedia</td>
<td>46</td>
<td>9,744</td>
<td>10,784</td>
<td>16,787</td>
<td>40,511</td>
<td>13,942</td>
<td>128,18</td>
<td>126,140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizations</td>
<td>172</td>
<td>15,559</td>
<td>14,358</td>
<td>187,842</td>
<td>431,232</td>
<td>44,139</td>
<td>947,26</td>
<td>1,139,170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in DBpedia</td>
<td>115</td>
<td>6,317</td>
<td>4,940</td>
<td>8,695</td>
<td>15,984</td>
<td>12,907</td>
<td>60,54</td>
<td>44,458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locations</td>
<td>85</td>
<td>10,356</td>
<td>8,339</td>
<td>40,859</td>
<td>24,729</td>
<td>23,894</td>
<td>290,05</td>
<td>228,445</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in DBpedia</td>
<td>81</td>
<td>7,773</td>
<td>7,369</td>
<td>11,364</td>
<td>16,372</td>
<td>11,167</td>
<td>88,65</td>
<td>76,341</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triples</td>
<td>95,219,534</td>
<td>105,675,519</td>
<td>110,861,823</td>
<td>316,034,616</td>
<td>240,731,408</td>
<td>188,296,316</td>
<td>535,035,57</td>
<td>1,240,774,944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from Mentions</td>
<td>1,046,544</td>
<td>9,700,585</td>
<td>16,888,833</td>
<td>46,359,300</td>
<td>136,135,841</td>
<td>65,631,222</td>
<td>439,060,64</td>
<td>1,146,601,954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from DBpedia</td>
<td>94,172,990</td>
<td>95,974,934</td>
<td>94,172,990</td>
<td>269,875,316</td>
<td>104,595,567</td>
<td>122,665,094</td>
<td>95,974,93</td>
<td>94,172,990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distilled from</td>
<td>DBpedia 2015</td>
<td>DBpedia 2015</td>
<td>DBpedia 2015</td>
<td>DBpedia 3.9</td>
<td>DBpedia 3.9</td>
<td>DBpedia 2014</td>
<td>DBpedia 201</td>
<td>DBpedia 201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• **Publications**
  
  
  
  
  - *Integrating Unstructured and Structured Knowledge with the KnowledgeStore* (Marco Rospocher, Francesco Corcoglioniti, Roldano Cattoni, Bernardo Magnini, Luciano Serafini) in Proceedings of the Posters and Demos of the 19th International Conference on Knowledge Engineering and Knowledge Management (EKAW2014).
  
  - *A Simple API to the KnowledgeStore* (Ian Hopkinson, Steve Maude, Marco Rospocher) in Proc. of ISWC Developers Workshop colocated with 13th Int. Semantic Web Conference (ISWC’14), Riva del Garda, Italy.

• **Deliverables**
  
  - **D6.1: KnowledgeStore Design** (June 2013)
  
  - **D6.2.1: KnowledgeStore version 1** (December 2013)
  
  - **D6.2.2: KnowledgeStore version 2** (February 2015)
  
  - **D6.2.3: KnowledgeStore version 3** (October 2015)

• **Demos**
  
  - Demonstration video showing accessing the KnowledgeStore through the UI - [https://youtu.be/YVOQajJt4](https://youtu.be/YVOQajJt4)

• **Publicly Accessible KnowledgeStore Instance**
  
  - Wikinews KnowledgeStore: [https://knowledgestore2.fbk.eu/nwr/wikinews/ui](https://knowledgestore2.fbk.eu/nwr/wikinews/ui)
KnowledgeStore

Scalable storage for text and RDF data

Overview

Despite the widespread diffusion of structured data sources and the public acclaim of the Linked Open Data initiative, a preponderant amount of information remains nowadays available only in unstructured form, both on the Web and within organizations. While different in form, structured and unstructured contents speak about the very same entities of the world, their properties and relations; still, frameworks for their seamless integration are lacking. The NewsReader KnowledgeStore is a scalable, fault-tolerant, and Semantic Web grounded storage system to jointly store, manage, retrieve, and semantically query, both structured and unstructured data. The KnowledgeStore plays a central role in the NewsReader EU project: it stores all contents that have to be processed and produced in order to extract knowledge from news, and it provides a shared data space through which NewsReader components cooperate.

Resources

- The KnowledgeStore API JavaDoc and browsable source Code
- Fork the KnowledgeStore through the Github page
Knowledge Store
— Scalable Framework for Interlinking Text and Knowledge —

The Team

Roldano Cattoni, Francesco Corcoglioniti, Bernardo Magnini, Alessio Palmero Aprosio, Mohammed Qwaider, Marco Rospocher, Luciano Serafini