Industrial implementations of IKS
Nuxeo

8 April 2011
Stefane Fermigier, Nuxeo
Nuxeo

- Founded in 2000, 50 employees
- Located in Paris, Boston and SF
- Open source stack rewritten in 2006, using open source Java technologies
  - Large code base (400 kSLOC of Java), very modular and extensible (thanks in part to OSGi)
- Main markets:
  - Document Management (DMS) -> public administrations and large companies, alternative to Documentum or OpenText
  - Repository / Platform -> custom / complex applications
  - News management system -> press agencies
  - Digital Asset Management (DAM) -> various
Nuxeo ECM - Our Approach

**Business Applications**
- Construction
- Media
- Government
- Life Sciences
- Correspondence Management
- Contracts Management
- Invoice Processing
- Records Management

**Packaged Products**
- Document Management
- Digital Asset Management
- Case Management Framework

**Foundation**
- Content Application Platform
- Core Server

**Technology**
- Nuxeo Enterprise Platform
  Complete set of components covering all aspects of ECM. Extensible modular architecture designed for content application development

**Support & Maintenance**
- Nuxeo Studio Marketplace
- Nuxeo Connect Subscription
Our (initial) expectations

<table>
<thead>
<tr>
<th>Improving DM and DAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLP (named entities, quotes, co-occurrences... extraction)</td>
</tr>
<tr>
<td>Pictures analysis</td>
</tr>
<tr>
<td>Document classification and clustering</td>
</tr>
<tr>
<td>Faceted semantic search</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correspondence / Case Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyze and route incoming mails / cases depending on their content</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaboration and KM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promote emergence, capture and sharing of knowledge within an organization by combining social and semantic technologies</td>
</tr>
</tbody>
</table>
What I wrote for last year’s review

“The semantic part is “easy” (using state of the art libraries or algorithms) …”
- Actually I was wrong, lots of work has been going on on the FISE / Stanbol project

“… the hardest part is actual integration (UI, user feedback, changing internal data structures when needed), even with a modular, service-oriented and event-driven architecture such as ours”
- But I will show you some progress on this part today
This year’s focus

- We worked (quite a bit) on the “easy” part (FISE then Stanbol)
- Later in 2010, we’ve conducted a first integration of FISE / Stanbol / IKS Alpha Stack into Nuxeo DM
  - Focus on named entities suggestion / management
  - Leverage Linked Open Data
- Iterative approach
From Tags to Entities

- Nuxeo already knows how to classify documents by tag or Dublin Core metadata (e.g. DC Subject)

- So what’s wrong with tags and metadata?
  - Several terms for same meaning
  - Several meanings for same terms
    - Need context to remove any ambiguity
Jimi Hendrix
James Marshall "Jimi" Hendrix (born Johnny Allen Hendrix; November 27, 1942 – September 18, 1970) was an American<br>guitarist, singer, songwriter and producer. He was known as the "King of Electric Guitar" and is often considered one of the greatest guitarists of all time.

Elmore James
Elmore James (January 27, 1918 – May 24, 1963) was an American blues guitarist, singer, songwriter and record producer. He was known for his unique style and his use of the phrase "May the Wind Blow Free".

Howlin' Wolf
Chester Arthur Burnett (June 10, 1910 – January 10, 1976), better known as Howlin' Wolf, was an influential American blues<br>singer, songwriter and musician. He is known for his distinctive voice and his contributions to the development of the blues genre.

Albert King
Albert King (April 25, 1923 – December 21, 1992) was an American blues guitarist and singer. He is known for his unique style and his use of the phrase "May the Wind Blow Free".

B.B. King
Riley B. King (born September 16, 1925), known by the stage name B.B. King, is an American blues guitarist and singer-songwriter.

Muddy Waters
McKinley Morganfield (April 4, 1913 – April 30, 1983), known as Muddy Waters, was an American blues musician, generally considered one of the greatest guitarists of all time.

Steve Cropper
Steven Lee Cropper (born October 21, 1941 in Willow Springs, Missouri), also known as Steve "The Colonel" Cropper is an American<br>guitarist.

American Red Cross
The American Red Cross (also known as the American National Red Cross) is a humanitarian organization that provides emergency assistance to those affected by disasters, diseases, and other public health crises.

Apple Inc.
Apple Inc. is an American multinational technology corporation that designs and manufactures consumer electronics, computer software, and consumer<br>products. It is headquartered in Cupertino, California.

Guardian Media Group
Not to be confused with the Guardian Enterprise Group, an American media company. Guardian Media Group plc (often referred to as GMG) is a British<br>media company.

International Red Cross and Red Crescent Movement
The International Red Cross and Red Crescent Movement is an international humanitarian movement with approximately 97 million<br>volunteers.

O'Reilly Media
O'Reilly Media (formerly O'Reilly & Associates) is an American media company established by Tim O'Reilly that publishes books and<br>other content.

Stanford University
The Leland Stanford Junior University, commonly referred to as Stanford University or Stanford, is a private research university in Palo Alto, California.

Starbucks
Starbucks Corporation is an international coffee and<br>coffeehouse chain based in Seattle, Washington, United States. Starbucks is the largest<br>coffeehouse chain in the world.

Woodstock, New York
Woodstock was a town in Ulster County, New York, United States. The population was 6,241 at the 2000 census. The Town of Woodstock is a<br>statutory town.
Washington is...

- **George Washington**: George Washington (February 22, 1732 – December 14, 1799) served as the first President of the United States from 1789 to 1797 and...}

- **George Washington University**: The George Washington University (GW, GWU, or George Washington) is a private, coeducational comprehensive research university located...]

- **Washington Redskins**: The Washington Redskins are a professional American football team based in the Washington, D.C. area. The team plays at FedExField[...]

- **Washington University in St. Louis**: Washington University in St. Louis is a nonsectarian, private research university located in suburban St. Louis, Missouri. Founded[...]

- **Washington**: Washington is a state in the Pacific Northwest region of the United States. Washington was carved out of the western part of Washi[...]

- **Washington, D.C.**: Washington, D.C. formally the District of Columbia and commonly referred to as Washington, the District, or simply D.C., is the[...]
Tagging with Entities

Linking with LOD brings:

- Global namespace / universal meaning context
- Interoperability across domains
- Interoperability across applications
Next steps for entity management

- Integration of the Entity Hub to ensure fast and reliable access to LOD
  - Note: most commercial NER vendors lack this part
- Current architecture can already be used to easily plug in commercial vendors tools
- Multilingual models
- New entity types
  - Generic: commercial products, artworks...
  - Specific: proteins, genes, drugs...
More engines

- Text summarization
- Topic classification
- Picture analysis
  - Face detection / recognition
  - Clustering
- Faceted semantic search
Business perspectives

- Semantic features to be provided as add-ons to existing products
- Demo / beta version of the plugin already available, but limited spontaneous adoption
- A few interested customers or prospects so far, hopefully first projects could start in 2011
- Potential partnerships with several semantic technology vendors (NER, sem. search) could bring significant business
# The IKS Consortium

<table>
<thead>
<tr>
<th>Project Lead and Coordination</th>
<th>Wernher Behrendt</th>
</tr>
</thead>
</table>
| *Salzburg Research* | *Salzburg Research Forschungsgesellschaft m.b.H.*  
Jakob Haringer Straße 5/3 | 5020 Salzburg, Austria  
T +43.662.2288-409 | F +43.662.2288-222  
wernher.behrendt@salzburgresearch.at  
[www.salzburgresearch.at](http://www.salzburgresearch.at)

<table>
<thead>
<tr>
<th>Deutsches Forschungsinstitut für Künstliche Intelligenz (DFKI)</th>
<th>Universität St. Gallen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Institute of Technology Management</em></td>
</tr>
<tr>
<td></td>
<td><em>University of St.Gallen</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consiglio Nationale delle Ricerche (CNR)</th>
<th>Software Quality Lab Universität Paderborn</th>
</tr>
</thead>
</table>
|                                       | *s-lab*  
*Universität Paderborn* |

<table>
<thead>
<tr>
<th>Software Research and Development Consultancy Ltd (SRDC)</th>
<th>Hochschule Furtwangen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>IHFU</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nuxeo Sa.</th>
<th>Alkacon Software GmbH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Alkacon</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TXT Polymedia</th>
<th>Pisano Holding GmbH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>PISANO</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nemein Oy</th>
<th>Day Software AG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Day</em></td>
</tr>
</tbody>
</table>