Partner presentation

Dr. José Manuel Gómez-Pérez, Esteban García Cuesta

xLiMe KickOff meeting
25th – 27th November 2013
“Our mission is to transform research into innovation and value for our clients in the networked economy”

http://www.isoco.com
What we do

User Communities

Suppliers

Information (a lot!)

Business

Clients

Partners

Competitors

Internet

Collaborative Network

Knowledge Network
Help customers increase productivity by creating and maintaining collaboration networks with clients, suppliers and partners.

OPEN INNOVATION
Help customers increase productivity by creating and maintaining collaboration networks with clients, suppliers and partners.

Enable intelligent, multimodal access to corporate and external information resources.
1. **OPEN INNOVATION**
   - Help customers increase productivity by creating and maintaining collaboration networks with clients, suppliers and partners.

2. **NATURAL INTERACTION**
   - Enable intelligent, multimodal access to corporate and external information resources.
   - Information (a lot!)

3. **BUSINESS INTELLIGENCE**
   - Support customers to adapt to changes in the market and optimize their resources.

- Open Ideas
- iQuotes
- iDesk
Help customers increase productivity by creating and maintaining collaboration networks with clients, suppliers and partners.

Enable intelligent, multimodal access to corporate and external information resources.

Support customers to adapt to changes in the market and optimize their resources.

Semanti-K, Open Ideas, iQuotes, iDesk

User Communities, Suppliers, Clients, Partners, Business, Competitors
Our customers

Public administration
- Cámara Madrid
- Ajuntament de Barcelona
- Fundación Marcelino Botín
- Generalitat de Catalunya
- Barcelona Activa
- Zaragoza

Industry, Energy and Telco
- Coca-Cola
- Telefónica
- Repsol YPF
- BT
- gasNatural
- affinity
- HP
- Yamaha

Finance and Insurance
- bankinter
- Deutsche Bank
- "la Caixa"
- Bekaert
- Liberty Seguros
- MAPFRE

Entertainment
- AERCEC
- everest
- PortAventura
- fnac
- Grupo Planeta

Health and Pharma
- Almirall
- ferr
- AstraZeneca
- Boehringer Ingelheim
- LABORATORIOS KIN
- CONSORCI HOSPITAL GENERAL UNIVERSITARI VALÈNCIA
- Hospital Universitario de Fuenlabrada
- Comunidad de Madrid

iSOCO
enabling the networked economy
Innovation culture

Research

Technology

Market
Dealing with the data deluge

A lifecycle for information management

Extraction

Structure

Access, Sharing, Reuse
Dealing with the data deluge

A lifecycle for information management

Research areas

- Knowledge Acquisition and Representation
- Data integration and interoperability
- Provenance
- Intelligent Information Access
- Information Retrieval
- Recommendation
- Visualization and Navigation
Identification of relevant information
- Reuse
- Data sources
- Quality
- Provenance

Information Extraction
- (Semi)structured data integration
- (Non-structured)
- Named-Entity Recognition
- Relation Extraction
- Clustering

Formalization
- Semantic annotation
- Disambiguation
- Classification
- Indexing

Dataset generation
- Publication in RDF as Linked Data
- Interlinking
- Encapsulation
- Co-creation
- Curation

Access
- Search
- Exploration
- Visualization
- Analytics
- Social
- Recommendation

A lifecycle for information management

Technological Pipeline

Reuse!
Semantic search engines for municipal procedures
Bienvenido a Endesa
¿En qué te podemos ayudar?

Yo: cambio de tarifa
Lucca: Tienes que entrar a la Oficina Online con tu usuario y clave y solicitar el cambio a través del formulario que encontrarás en la opción "Modificación de contrato" - Datos técnicos. La modificación no se realiza inmediatamente.

Introduce tu pregunta: [Entrada]
Enviar

Que nada te ate, tu energía va contigo
Descarga la nueva App Endesa y disfruta de todas sus ventajas

GESTIONES ONLINE
Nuestra Oficina Online
Nuestras Apps
Factura Electrónica

¿Qué necesitas?
Luz
Comienza a ahorrar en tus facturas de la luz con nuestras tarifas.
Tarifa Luz Endesa Online

Gas
Las tarifas de Gas que mejor se adaptan a tus necesidades de consumo.
Tarifa Gas Endesa Online

Te interesas
Te asesoramos

MEJOR QUE UN TRIPLE EN EL ÚLTIMO SEGUNDO?
Gratuitas entradas a este, balones en el sorteo de 10 años de luz.

¿Qué necesitas?
1 2 3 4
Hi Sarah! Following croaks might be of your interest since they are about:

Ontology instances: "offer" "krow"

THURSDAY, 17 MARCH 2011

KROW demo is ready for public release. Fingers crossed!
By William 90 minutes ago @Add Comment

Great work from the KROW team! Spread the word people!
By Peter 52 minutes ago @Add Comment

Fixing some last minute bugs in the KROW demo
By Paul 1 hour, 9 minutes ago @Add Comment

BEST IN-USE PAPER ESWC'11
Knowledge Tagger enables users to perform Named Entity Resolution in texts using relevant domain ontologies and semantic data. Its distinguishing characteristic is its customization capabilities as it allows users to define and apply Ontology-Based Disambiguation Evidence Models, based on their knowledge about the domain(s) and expected content of the texts to be analyzed.

To construct an evidence model the user needs to do only two things: i) determine the concepts whose instances want to detect/disambiguate and ii) determine the related to them concepts whose instances may serve as contextual disambiguation evidence within the expected texts. Having done that, he/she may automatically generate a disambiguation model and immediately apply it to relevant texts.

As an example, we have already defined a model appropriate for the disambiguation of football related entities within texts that describe highlights of football matches. The model may be tested by using the form below and analyzing any of the given example texts. Moreover, new evidence models may be created by pressing the relevant button and filling in the relevant form. As soon as a model is created, it becomes available for use. In the current version of this demo new models may be defined only for the History domain, but more will be added soon.

An exemplar scenario that may be used to test the new model creation functionality may be found here. More detailed information about the system’s underlying framework is also available here.

**Defined Evidence Models:**

1. Teams and Players in Football Match Descriptions of the Spanish Liga

<table>
<thead>
<tr>
<th>Target Concept</th>
<th>Evidence Concept</th>
<th>Relation(s) Linking Evidence to Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:SoccerPlayer</td>
<td>dbpedia-owl:SoccerClub</td>
<td>is dbpprop:currentClub of</td>
</tr>
</tbody>
</table>
Knowledge Tagger enables its users to perform Named Entity Resolution in texts using relevant domain ontologies and semantic data. Its distinguishing characteristic is its customization capabilities as it allows users to define and apply Ontology-Based Disambiguation Evidence Models, based on their knowledge about the domain(s) and expected content of the texts to be analyzed.

To construct an evidence model, the user needs to do only some basic steps: selecting the related to them concepts whose instances may serve as a basis for evidence and automatically generate a disambiguation model and immediate answers.

As an example, we have already defined a model appropriate to sports and football matches. The model may be tested by using the form below, by pressing the relevant button and filling in the relevant form. Note that an evidence model may be defined only for the History domain, but not for the newest sport.

An exemplar scenario that may be used to test the new model and its underlying framework is also available here.

Defined Evidence Models:

1. **Teams and Players in Football Match Descriptions of the Spanish Liga**

<table>
<thead>
<tr>
<th>Target Concept</th>
<th>Evidence Concept</th>
<th>Relation(s) linking Evidence to Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia-owl:SoccerClub</td>
<td>dbpedia-owl:SoccerPlayer</td>
<td>is.dbpprop:currentclub.of</td>
</tr>
<tr>
<td>dbpedia-owl:SoccerClub</td>
<td>dbpedia-owl:SoccerClub</td>
<td>is.dbpprop:currentclub.of</td>
</tr>
<tr>
<td>dbpedia-owl:SoccerClub</td>
<td>dbpedia-owl:SoccerPlayer</td>
<td>dbpprop:currentclub, is.dbpprop:currentclub.of</td>
</tr>
<tr>
<td>dbpedia-owl:SoccerClub</td>
<td>dbpedia-owl:SoccerManager</td>
<td>dbpprop:manager</td>
</tr>
</tbody>
</table>

Examples:

- This lost penalty kick by Llorente cost Athletic the cup, I think. To be honest, I am **not** disappointed right now. 
- It is the 70th minute of the game and after a magnificent pass by Pedro, Messi, managed to beat Claudio Bravo. Barcelona now leads 1-0 against Real Madrid.
- **Pedro Rodriguez Ledesma** 0.38
- **Pedro Mario Alcaraz** 0.23
- **Pedro Bigas** 0.10
- **Pedro Lopes Miguez** 0.10
- **Pedro Rados** 0.10
- **Pedro Roberto Silva Rotinha** 0.10

Input text:

It is the 70th minute of the game and after a magnificent pass by Pedro, Messi, managed to beat Claudio Bravo. Barcelona now leads 1-0 against Real Madrid.
Data integration and analytics
- Customers in touristic areas who are not buying Sprite
- Introduce a new KPI in less than 3 months
- Customers demanding “Nordic Mist” but not "Coca-Cola Zero"
- Bars with Pay TV
Research Object

Content aggregation
Research Objects

http://www_wf4ever-project.org

Scientists

Electronic paper

Hypothesis

Annotations

Datasets

Results

Provenance

Workflows

Experiments

Workflows
Research Object

- Carriers of Research Context
- Uniquely identifiable and referentiable
- Metadata
  - Annotations, Manifests, Recipes, Permissions, Discourse
  - Provenance
- Lifecycle
  - Evolution, versioning
- Mixed Stewardship
  - Distributed, dispersed, local and external, 3rd party tenancy, heterogeneous
  - Decay
  - Graceful Degradation

RO Primer: [http://wf4ever.github.com/ro-primer](http://wf4ever.github.com/ro-primer)
RO specification: [http://wf4ever.github.com/ro](http://wf4ever.github.com/ro)
DATA REUSE
AGGREGATION
QUALITY
ACCESS
Monitoring quality against decay

The RO monitoring tool

http://sandbox.wf4ever-project.org/decayMonitoring/monitor.html

BEST PAPER
SEPUBILCA’13
CollabSpheres (exploratory + discovery)

Advanced search interfaces

CollabSpheres v.0.1

User
Collaborators
Context of interest
Contextual recommendation
ROs
Historic recommendation

Information about a selected resource.

Application example: Media-Learning

**Information objects**

- **News as content aggregation (information objects)**
  - Entities capable to change, evolve, and connect with other resources and users
  - Timeline

- **Added value for end users**
  - Social
  - Recommendation and personalized content
  - Extended with related information
  - Exploratory search

- **Traceability**
  - Where does the news come from?
  - Can I trust it?
  - Is it up to date?

- **Quality monitoring**
  - Is the information object fit?
Research & Innovation projects and Partners

Enablers
Main involvement in the project

**WT6:**

**Project Effort by Beneficiary and Work Package**

<table>
<thead>
<tr>
<th>Beneficiary number and short-name</th>
<th>WP 1</th>
<th>WP 2</th>
<th>WP 3</th>
<th>WP 4</th>
<th>WP 5</th>
<th>WP 6</th>
<th>WP 7</th>
<th>WP 8</th>
<th>WP 9</th>
<th>Total per Beneficiary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - KIT</td>
<td>4.00</td>
<td>12.00</td>
<td>8.00</td>
<td>24.00</td>
<td>13.00</td>
<td>2.00</td>
<td>2.00</td>
<td>7.00</td>
<td>21.00</td>
<td>93.00</td>
</tr>
<tr>
<td>2 - JSI</td>
<td>6.00</td>
<td>16.00</td>
<td>6.00</td>
<td>14.00</td>
<td>13.00</td>
<td>2.00</td>
<td>2.00</td>
<td>7.00</td>
<td>0.50</td>
<td>66.50</td>
</tr>
<tr>
<td>3 - UNITN</td>
<td>4.00</td>
<td>16.00</td>
<td>32.00</td>
<td>9.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>8.00</td>
<td>0.50</td>
<td>71.50</td>
</tr>
<tr>
<td>4 - ISOCO SA</td>
<td>10.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.00</td>
<td>38.00</td>
<td>10.00</td>
<td>5.00</td>
<td>1.00</td>
<td>73.00</td>
</tr>
<tr>
<td>5 - ZATTOO</td>
<td>10.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.00</td>
<td>2.00</td>
<td>12.00</td>
<td>4.00</td>
<td>0.50</td>
<td>30.50</td>
</tr>
<tr>
<td>6 - VICO RESEARCH &amp; CONS</td>
<td>10.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>2.00</td>
<td>2.00</td>
<td>14.00</td>
<td>5.00</td>
<td>0.50</td>
<td>35.50</td>
</tr>
<tr>
<td>7 - econda</td>
<td>6.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>4.00</td>
<td>2.00</td>
<td>16.00</td>
<td>5.00</td>
<td>0.50</td>
<td>33.50</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50.00</td>
<td>44.00</td>
<td>48.00</td>
<td>47.00</td>
<td>43.00</td>
<td>49.00</td>
<td>57.00</td>
<td>41.00</td>
<td>24.50</td>
<td>403.50</td>
</tr>
</tbody>
</table>

- Main effort in WP6 (integration & toolkit)
- Also, significant effort in other WPs, especially in
  - WP1 (processing multilingual, multimedia data)
  - WP5 (cross-media, cross-lingual analytics)
  - WP7 (use cases & evaluation)
» Esteban García (project manager)
» Nuria García (software engineer)
» Aleix Garrido (software engineer)
» Almudena González (software engineer)
» José Manuel Gómez (director R&D)
Previous Work
xLiMe Integration

![Diagram showing integration of languages, visual, auditiv, and textual inputs with cross-lingual annotations.](image)

- **Languages**: TV, Video Clips, Photos, Images; Social Video, Social Photos; Podcasts, Audio from Video; News, Descriptions of video/audio; Twitter, Blogs, Comments, Tags
- **Visual**: TV, Video Clips, Photos, Images
- **Auditiv**: Podcasts, Audio from Video
- **Textual**: News, Descriptions of video/audio; Twitter, Blogs, Comments, Tags
- **Cross-lingual annotations**: Integrated, disambiguated, cross-lingual, cross-media, knowledge base

**iSoco**

enabling the networked economy
xLiMe Integration
The Xlike pipeline provides an easy incremental way of transferring functionalities from each work package to the next ones.
... by functionality

**HUMAN COMPUTING INTERFACE (HCI)**
- Trends and Analytics
- Mapping and Geolocalization

**ANALYSIS & INTERPRETATION**
- Search
- Categorization
- Story identification
- Language Identification
- Text Annotation
- Shallow Linguistic Processing
- Deep Linguistic Processing
- Cross-Lingual Text Annotation

**ACQUISITION**
- JSI NewsFeed
- Zattoo
- Vico
- Econda

xLiMe Integration
Sandbox Architecture
xLiMe Integration

Sandbox Architecture

---

**ACQUISITION**

JSI NewsFeed

**ACQUISITION**

Zattoo

Econda

---

**HUMAN COMPUTING INTERFACE (HCI)**

Trends and Analytics

Mapping and Geolocation

---

**ANALYSIS & INTERPRETATION**

Language Identification

Text Annotation

Shallow Linguistic Processing

Deep Linguistic Processing

Cross-Lingual Text Annotation

---

**SANDBOX**

Search

Categorization

Story identification

---

... by functionality

---
xLiMe Integration

Sandbox Architecture

... by functionality

HUMAN COMPUTING INTERFACE (HCI)

Trends and Analytics
Mapping and Geolocalization

ANALYSIS & INTERPRETATION

Search
Categorization
Story identification

Language Identification
Text Annotation
Cross-Lingual Video Annotation
Cross-Lingual Audio Annotation
Cross-Lingual Text Annotation

ACQUISITION

JSI NewsFeed

ACQUISITION

Zattoo
Vico
Econda

Multilingual data feeds
Multilingual Linguistic processing
Cross-lingual Semantic Annotation
Reporting & Analytics components
Toolkit & Interface

SANDBOX

Shallow Linguistic Processing
Deep Linguistic Processing

Mapping and Geolocalization

Cross-Lingual Text Annotation

Econda

Vico
ACQUISITION

ANALYSIS & INTERPRETATION

HUMAN COMPUTING INTERFACE (HCI)

Trends and Analytics
Mapping and Geolocalization

SEARCH
Categorization
Story identification

Multilingual Analysis
Cross-Lingual Video Annotation
Cross-Lingual Audio Annotation
Cross-Lingual Text Annotation

ACQUISITION

JSI NewsFeed

SANDBOX

… by functionality

xLiMe Integration
Sandbox Architecture

Reporting & Analytics components
Cross-lingual Semantic Annotation
Multi-lingual Semantic Annotation
data feeds

Multilingual Linguistic processing

Mapping and Geolocalization

Shallow Linguistic Processing
Deep Linguistic Processing

Language Identification

JSI NewsFeed
Zattoo
Vico
Econda

SANDBOX
Linguistic Analysis
- Language Identification
- Tokenization
- POS Tagging
- Named Entity Recognition

Semantic Annotation
- Named Entity Annotation

Cross-lingual Linking
- Named Entity Linking

SandBox XLike

MultiLingual Language Processing

Examples:
- Bruce Springsteen is an American singer-songwriter and multi-instrumentalist.
- Blade Runner es una película de ciencia ficción estadounidense dirigida por Ridley Scott.
- LiPhone es un dispositivo electrónico multimédia presentat per Apple Computer el 9 de gener de 2007.
- Der russische Schachspieler Garri Kasparow wurde 1985 proklamierte die jüngste Weltmeister in der Geschichte.
- 正面的观点认为，由于王朝内部权力的分化就开始“行灭亡”.
- Clara je bila žena skladatelja Roberta Schumanna in ena vodilnih pianistov in skladateljev romantike.

Select language:
- Automatic
- English
- Spanish
- Catalan
- German
- Chinese
- Slovene

[Analyze]
xLiMe Integration

XLike Sandbox

MultiLingual Language Processing

Examples:
- Bruce Springsteen is an American singer-songwriter and multi-instrumentalist.
- El Último es un compositor de música folclórica español dirigido por Ridley Scott.
- L’iPhone è un dispositivo elettronico multimediale presentato per Apple Computer il 9 di gennaio 2007.
- Der russische Schauspieler Gorri Kaspauer wurde 1985 proklamiert als „Großartigster Schauspieler der Geschichte“.
- Чир як було сказано Роберта Шумана в нові переклади та в традиціях романики

Bruce Springsteen is an American singer-songwriter and multi-instrumentalist.

TC-04 Multilingual Analysis

TC-03 Language Identification

TC-08 Named Entity Annotation
xLiMe Integration

XLike Toolkit

Requirements – Components Relation

<table>
<thead>
<tr>
<th>Requirement Identifier</th>
<th>Component Identifier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC1</td>
<td>TC-01</td>
<td>NewsFeed</td>
</tr>
<tr>
<td></td>
<td>TC-03</td>
<td>Language Identification</td>
</tr>
<tr>
<td></td>
<td>TC-04</td>
<td>Multilingual Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-05</td>
<td>Deep Linguistic Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-08</td>
<td>Name Entity Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-09</td>
<td>Wikipedia Miner WikiFiler Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-10</td>
<td>Early Ontological Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-11</td>
<td>Crowd Sourcing Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-12</td>
<td>Finale Ontological Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-13</td>
<td>Final Text Annotation Service</td>
</tr>
<tr>
<td></td>
<td>TC-14</td>
<td>Early Machine Translation based Semantic Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-15</td>
<td>Final Machine Translation based Semantic Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-16</td>
<td>Cross-lingual UFP</td>
</tr>
<tr>
<td></td>
<td>TC-17</td>
<td>KIT Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-18</td>
<td>JSI Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-19</td>
<td>Cross-lingual Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-20</td>
<td>Cross-lingual Document Linking</td>
</tr>
<tr>
<td></td>
<td>TC-25</td>
<td>News Data Visualization Component</td>
</tr>
</tbody>
</table>

| UC3                    | TC-01                | NewsFeed    |
|                        | TC-06                | Linguistic Relation Extraction |
|                        | TC-07                | Informal Language Analysis |
|                        | TC-17                | KIT Cross-lingual Similarity |
|                        | TC-18                | JSI Cross-lingual Similarity |
|                        | TC-19                | Cross-lingual Analysis |
|                        | TC-20                | Cross-lingual Document Linking |
|                        | TC-25                | News Data Visualization Component |

Lightweight RESTful approach

Toolkit Deployment Diagram

- News Data Visualization TC-23
- Multilingual Analysis (zh) TC-04
- Real-time S/A TC-02
- Language Identification TC-03
- Multilingual Analysis (en, es, zh) TC-04
- Multilingual Analysis (zh) TC-04
- Newsfeed TC-01
- Multilingual Analysis (zh) TC-04
- Cross-lingual Similarity TC-16
- Event Extraction (new) TC-20
- Detection of news reporting bias (new) TC-21
### Requirements – Components Relation

**Good for tracking, for us and for reviewers**

#### Toolkit Deployment Diagram

[Diagram showing components and their relations]

<table>
<thead>
<tr>
<th>Requirement Identifier</th>
<th>Component Identifier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC1</td>
<td>TC-01, TC-05</td>
<td>NewsFeed</td>
</tr>
<tr>
<td></td>
<td>TC-03</td>
<td>Language Identification</td>
</tr>
<tr>
<td></td>
<td>TC-04</td>
<td>Multilingual Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-05</td>
<td>Deep Linguistic Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-08</td>
<td>Name Entity Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-09</td>
<td>Wikipedia Miner WikiRef Annotation</td>
</tr>
<tr>
<td></td>
<td>TC-10</td>
<td>Early Ontological Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-11</td>
<td>Crowd Sourcing Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-12</td>
<td>Final Ontological Word-sense-disambiguation</td>
</tr>
<tr>
<td></td>
<td>TC-13</td>
<td>Final Text Annotation Service</td>
</tr>
<tr>
<td></td>
<td>TC-14</td>
<td>Early Machine Translation based Semantic Ar</td>
</tr>
<tr>
<td></td>
<td>TC-15</td>
<td>Final Machine Translation based Semantic Ar</td>
</tr>
<tr>
<td></td>
<td>TC-16</td>
<td>Cross-lingual USP</td>
</tr>
<tr>
<td></td>
<td>TC-17</td>
<td>KIT Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-18</td>
<td>JSI Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-19</td>
<td>Cross-lingual Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-20</td>
<td>Cross-lingual Document Linking</td>
</tr>
<tr>
<td></td>
<td>TC-25</td>
<td>News Data Visualization Component</td>
</tr>
<tr>
<td>UC2</td>
<td>TC-01</td>
<td>NewsFeed</td>
</tr>
<tr>
<td></td>
<td>TC-06</td>
<td>Linguistic Relation Extraction</td>
</tr>
<tr>
<td></td>
<td>TC-07</td>
<td>Informal Language Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-17</td>
<td>KIT Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-18</td>
<td>JSI Cross-lingual Similarity</td>
</tr>
<tr>
<td></td>
<td>TC-19</td>
<td>Cross-lingual Analysis</td>
</tr>
<tr>
<td></td>
<td>TC-20</td>
<td>Cross-lingual Document Linking</td>
</tr>
<tr>
<td></td>
<td>TC-25</td>
<td>News Data Visualization Component</td>
</tr>
<tr>
<td>UC3</td>
<td>TC-01</td>
<td>NewsFeed</td>
</tr>
<tr>
<td></td>
<td>TC-06</td>
<td>Linguistic Relation Extraction</td>
</tr>
</tbody>
</table>
Data Centric Approach
Yesterday a house was consumed by fire between Henderson and Belle Plaine in Sibley County.
Previous Work

Schema and Data Model

<nodes>
  <node type="entity" class="organization" displayName="assistant_fire_chief" id="E10">
    <mentions>
      <mention sentenceId="6" id="E10.1" words="Assistant Fire Chief">
        <mention_token id="6.3"/>
      </mention>
    </mentions>
  </node>
  <node type="entity" class="location" displayName="belle_plaine_fire_departments" id="E2">
    <mentions>
      <mention sentenceId="1" id="E2.1" words="Belle Plaine">
        <mention_token id="1.11"/>
      </mention>
      <mention sentenceId="5" id="E2.2" words="Belle Plaine">
        <mention_token id="5.1"/>
      </mention>
      <mention sentenceId="6" id="E2.3" words="Belle Plaine Fire Departments">
        <mention_token id="6.6"/>
      </mention>
    </mentions>
  </node>
</nodes>
It does not have to be the same schema but the sooner the better
Time Performance

log (Time (ms.))

log (Number of words)

English
German
Spanish
Apache Cassandra

- Apache Cassandra is an open source distributed database management system designed to handle very large amounts of data spread out across many commodity servers while providing a highly available service with no single point of failure
- No-SQL.
- CQL is a SQL (Structured Query Language)-like language for querying Cassandra.
- CQL doesn't support joins, which make no sense in Cassandra.

- Rough study based on the current newsfeed data volume
  - 228MB generated per day and 81GB per year (before adding xlike metadata)
  - 1809MB generated per day and 644GB per year (after adding xlike metadata)
We are migrating to Cassandra in XLike architecture as storage service for allowing an easy way to store and access to already captured data.
We are migrating to Cassandra in XLike architecture as storage service for allowing an easy way to store and access to already captured data.
Apache Cassandra

- High performance in data access

- Access to trends of entities on real time (we have it locally so far)
  - http://172.16.0.80:9191/DataSet/rest/date?q=2013-04-22%2010:00:00%20CEST
  - http://172.16.0.80:9191/TrendingTopics/rest/trendings
Can Chinese art be cutting edge?: Four months before the Tiananmen crackdown in 1989, a woman fired a gun in China's National Art Gallery -- all in the name of art.
xLiMe Integration
Visualization and Analytics

Stories Visualization and Searching Tool
Y1 Prototype

Editors News Recommender Mockup
Y2 Use case
Event: something that is happening in the world and it was not occurring before (e.g. Sandy hurricane October 22, 2012).

Def: anything that takes place or happens, especially something important; happening; incident
Event: something that is happening in the world and it was not occurring before (e.g. Sandy hurricane October 22, 2012).

- **Def**: anything that takes place or happens, especially something important; happening; incident

- **Def**: an occurrence regarded as a bare instant of space-time as contrasted with an object which fills space and has endurance
Event: something that is happening in the world and it was not occurring before (e.g. Sandy hurricane October 22, 2012).

Def: **anything that takes place or happens**, especially something important; happening; incident

Def: an occurrence regarded as a bare instant of space-time as contrasted with an object which fills space and has endurance

Def: Physics a single occurrence of a process, e.g. the ionization of one atom.

http://oxforddictionaries.com
Which is the rarest?

"Be"

"House"
How rare is it?

- the
- be
- untouchable
- nick_triggle
- iran
- earthquake
- drug; lord;
- same-sex; marriage;

Less to more rare [0-1]

Observations Expected
How rare is it?

**BBC News - Correspondents - Nick Triggle**
www.bbc.co.uk/news/correspondents/nicktriggle
BBC health correspondent Nick Triggle’s reflections and analysis on what is ... Throughout the 1800s it hovered around the 40 years of age mark in the UK, but ...

**BBC News - Is the NHS going to blow a gasket?**
www.bbc.co.uk/news/health-22607986
23/05/2013 - Nick Triggle Article written by Nick Triggle Health .... 06:46 UK time, Wednesday, 21 August 2013. There are two ways of looking at the ...

**BBC News - Pledge to close health and care gap**
www.bbc.co.uk/news/health-22515978
13/05/2013 - Ministers are promising an end to the era of vulnerable people being passed around the health and care systems.

**BBC News - Five ways the NHS is changing**
www.bbc.co.uk/news/health-21964561
03/04/2013 - Nick Triggle Article written by Nick Triggle Health .... 00:46 UK time, Wednesday, 21 August 2013. There are two ways of looking at the ...

**BBC.co.uk_search_prweek.com_PRWeek UK__PRWeek_search**
www.prweek.com/uk/search/.../BBC.co.uk/.../rail
19/03/2010 - Key target media including the BBC.co.uk. The Press ... health editor - Richard Warny Richard.warny@bbc.co.uk Senior reporter - Nick Triggle ...

**Nick Triggle | BBC | ZoomInfo.com**
www.zoominfo.com/p/Nick-Triggle/494038624
How rare is it?

**BBC News - Correspondents - Nick Triggle**
www.bbc.co.uk/news/correspondents/nicktriggle

BBC health correspondent Nick Triggle’s reflections and analysis on what is... Throughout the 1800s it hovered around the 40 years of age mark in the UK, but...

**BBC News - Is the NHS going to blow a gasket?**
www.bbc.co.uk/news/health-22607986

23/05/2013 - Nick Triggle Article written by Nick Triggle Health .... 06:46 UK time, Wednesday, 21 August 2013. There are two ways of looking at the...

**BBC News - Pledge to close health and care gap**
www.bbc.co.uk/news/health-22515978

13/05/2013 - Ministers are promising an end to the era of vulnerable people being passed around the health and care systems.

**BBC News - Five ways the NHS is changing**
www.bbc.co.uk/news/health-21964561

13/05/2013 - Nick Triggle Article written by Nick Triggle Health .... 06:46 UK time, Wednesday, 21 August 2013. There are two ways of looking at the...

**PR Week UK - PRWeek search...**
www.prweekuk.com/search

20/05/2013 - Nick Triggle Article written by Nick Triggle Health .... 06:46 UK time, Wednesday, 21 August 2013. There are two ways of looking at the...

News Outbreak Detection
Visualization and Analytics

xlime

Emergent Interest

Viewer rates a show

And receives recommendations
Y1 Deliverables
Introduction

WP6: Integration and Toolkit

● Objectives:

● Implement a common lightweight integration platform for scalable and real-time execution of modules developed in previous work packages.

● Develop an API exposing functionality of the integration platform to the tools for exploratory real-time data feed analysis.

● Develop a Web front-end by exposing the integrated platform functionality through innovative visualization techniques.

● Validations: the ease of use of the integration of the xLime system in the environment of end users.
WP6: Integration and Toolkit

T6.1 Toolkit architecture (iSOCO, all) (M1 – M12)

T6.2 Integration platform (iSOCO, KIT, JSI, UNITN) (M4-36)

T6.3 API’s (iSOCO, KIT, JSI, UNITN) (M13-M24)

T6.4 Web front-end (iSOCO, KIT, JSI) (M19-M36)
• T6.1 Toolkit architecture (iSOCO, all) (M1 – M12)
• T6.2 Integration platform (iSOCO, KIT, JSI, UNITN) (M4-36)
• T6.3 API’s (iSOCO, KIT, JSI, UNITN) (M13-M24)
• T6.4 Web front-end (iSOCO, KIT, JSI) (M19-M36)
Deliverables

• **T6.1 Toolkit architecture**
  - **D6.1 Toolkit Architecture Specifications**: this deliverable will provide xLiMe toolkit architecture specifications updated based on the feedback from the rapid development cycle. Furthermore, an updated study of the real-time state-of-the-art tools will be presented.

• **T6.2 Integration Platform**
  - **D6.2.1 Early Prototype**: this deliverable will provide the early prototype of the integration platform. Here, the architecture specifications from D1.2 will be implemented and the modules required by D1.41 will be integrated. The robustness and independence of the different modules integrated in this integration platform will be demonstrated by appropriate tests and diagrams.
Deliverables

• T6.1 Toolkit architecture
  • **D6.1 Toolkit Architecture Specifications**: this deliverable will provide xLiMe toolkit architecture specifications updated based on the feedback from the rapid development cycle. Furthermore, an updated study of the real-time state-of-the-art tools will be presented.

• T6.2 Integration Platform
  • **D6.2.1 Early Prototype**: this deliverable will provide the early prototype of the integration platform. Here, the architecture specifications from D1.2 will be implemented and the modules required by D1.41 will be integrated. The robustness and independence of the different modules integrated in this integration platform will be demonstrated by appropriate tests and diagrams.

---

**Side Note: Beware!**

**Internal vs. Official Deadlines**

**ah a**

**Technology vs. Report Deadlines**
Objectives

WP6: Integration and Toolkit

Three different phases during the project:

1. **Requirements identification and specification**
   - Description of requirements
   - Description of architecture design
   - Integration model and components

2. **Construction of prototype and a scalable integration platform**
   - Execution of the pipeline from WP2-WP5
   - Scalability by using multiple threads/machines (Amazon EC2)

3. **Validation criteria**
   - Definition of use case scenarios
   - Expected results and their validation for real cases
WP6: Integration and Toolkit

Objectives

Establish an effective methodology for:

- Identify at **early stages** the **system requirements** from partners and end-users of the consortium
- Integrate the different **technological components** using a homogeneous perspective
- **Validation** of the components in the common platform
One important goal of WP6 is to provide the basis for end-user applications within WP7 and WP8 by providing the integration of a wide spectrum of multimedia, linguistic and analytics tools.

- **Support the developments** at different WPs of the project
- **Deploy and maintain a Sandbox**
- **Definition and design of the toolkit architecture and integration platform** driven by:
  - technical functionalities and available components, and use cases requirements
  - multi-content analysis pipelines and data flows
- **Explore/study the opportunities for scaling** the xLike architecture and technologies
- **Provide continuous development and integration of xLime implementation**
Next steps

Methodology

WP1

WP2, WP3 and WP4

WP5

WP6

Platform

Data Collection

Analysis

Interpretation

Integration

Front-End
Methodology

Next steps

WP1

WP2, WP3 and WP4

WP5

WP6

Platform

Data Collection

Analysis

Interpretation

Integration

Front-End
Howto

- The architecture is open to discussion
- It must be guided by technological partners, end-users, computational requirements, and using the experience in other technological projects
- How components will be integrated
  - Common framework
  - Service-oriented approach
- How algorithms will be adapted to the architecture
  - Software paradigms
  - Hardware needs
Understand which are the necessities and interactions in the project

- Early requirements gathering
  - Partner one to one meetings
  - Brainstorming and review
  - Agreement of requirements

- Initial definition of use cases
  - Initial set of requirements needed to accomplish with that use cases
  - Initial definition of components
  - Initial definition of interactions

- Finding out possible bottlenecks which could perturb the project
Co-evolutionary integration

• **Agile Methodology Development**
  - Small groups (2-3 people) with specific goals
  - Hackathons

• **Objectives**
  - Reference implementation of services (APIs)
  - Adheres to xLime scalable architecture (Sandbox)
  - Use case domains and HCI
  - Real time environments (scalability issues)
  - Towards exploitation and toolkit preservation

• **Architecture principles**
  - Distributed set of web services (SaaS architecture)
  - Loose coupling and independency by using REST interfaces (WP1, WP2, WP3, WP4, and WP5)
- Common sense
- Meetings
- Data driven
- Try to avoid that integration becomes a burden
Dr. Jose Manuel Gomez-Perez
Director, R&D
jmgomez@isoco.com

¿Quieres innovar?

Barcelona
Tel +34 935 677 200
Edificio Testa A
C/ Alcalde Barnils, 64-68
St. Cugat del Vallès
08174 Barcelona

Madrid
Tel +34 913 349 797
Av. del Partenón, 16-18, 1º7ª
Campo de las Naciones
28042 Madrid

Pamplona
Tel +34 948 102 408
Parque Tomás
Caballero, 2, 6º-4ª
31006 Pamplona

Valencia
Tel +34 963 467 143
Oficina 107
C/ Prof. Beltrán Bâguena, 4
46009 Valencia
In 2010 the size of the digital universe exceeded 1 Zettabyte (=1 trillion Gb)

» 1.8 Zb in 2011

» 35 Zb expected in 2020

» 90% unstructured data

» 70% user-generated

» 75% resulting from data copying, merging, and transforming

» Metadata is the fastest growing data category

» Much of such data is dynamic, real-time, volatile
A Pragmatic Approach to the Semantic Web

The Web of Data

The Linked Data Design Principles (Tim Berners-Lee, 2006)

1. URIs as unique identifiers
   › Anything, not just documents
2. Access the data through the HTTP web protocol
3. Provide the semantics of the data identified by those URIs
4. Interlink data with other data through RDF relations
   › Serendipity

Image by:
http://www.flickr.com/photos/PhOtOnQuAnTiQuE/
Main pillars

- Semantic search
- Linked Open Data
- Microblogging
- Social networks
- Context

Social micro-blogging

Twitter

Identi.ca
Collaboration

Interoperability and Access to Information

Web 1.0

Corporate Knowledge Management

http://mikrow.isoco.net
Knowledge Management

Corporate

Collaboration

Interoperability and Access to Information

Web 1.0

Web 2.0

http://mikrow.isoco.net
Knowledge Management

Corporate

Interoperability and Access to Information

Web 1.0

Web 2.0

Semantic Web

Collaboration

http://mikrow.isoco.net
Knowledge Management

Web 1.0

Web 2.0

Web 3.0

Semantic Web

Interoperability and Access to Information

Collaboration

http://mikrow.isoco.net
Enterprise 3.0: Collaborative aspects of Web 2.0 + Semantic Technologies applied to the enterprise
TransporteLinkedData.es

The INNPRONTA project CIUDAD 2020 aims to achieve significant advances in the areas of energy efficiency, internet of the future (IoF), internet of the things (IoT), human behavior, environmental sustainability and mobility & transport in order to design the city of the future: sustainable, smart and efficient.

CIUDAD 2020 conceives, designs and implements a new paradigm of smart, sustainable and efficient city around three fundamental axes: Energy and Efficiency, Mobility and Transport, and Environmental Control. The Ontology Engineering Group (OEG) is involved as a participant in the project by publishing data related to these three axes. Published sources are made available as RDF (Resource Description Framework) knowledge bases according to the Linked Data principles.

Case Study: Bicycle Sharing Systems

Bicycle sharing systems in different cities all over the world have made their data available in the Web. Bike rental stations are distributed in different points in the city and a system usually allows the user to pick up a bike at any station and drop it off at any (other) station. The goal of this case study is to publish up-to-date linked data about the availability of bikes and free slots in the stations of the different systems, and links to related resources like travel guides and points of interest, e.g. museums, restaurants.

Currently there is an API, http://api.cebik.es/, that allows the deployment of data about each sharing system either in JSON or in HTML format. Wrappers have been developed to extract the data of each system (Project: Pybikes). Examples of applications include the displaying the information on Android, and the generation of Javascript code that can be used to render maps in a Web page. However, by exploiting the benefits of linked data, applications can be enhanced and users may explore not only static and dynamic data of bike sharing systems, but also related (geo)linked information.

This work is being funded by the Centro para el Desarrollo Tecnológico Industrial as part of the INNPRONTA Spanish National Research Program.