Data Analytics Involving Text

Dunja Mladenić

Jožef Stefan Institute and 
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia
Outline

• Motivation
• Big Data
• Analysis of Text Data
• Conclusions

Our intelligence, our sophistication, is the key to our living!...
Old age without wisdom, youth without success and childhood without smiles are worthless.    [Bhajan, 2001]
Jožef Stefan Institute, Artificial Intelligence Laboratory

Jožef Stefan Institute (JSI) is the leading Slovene research institution for natural sciences (900+ people) in the areas of computer science, physics, chemistry, ecology.

Artificial Intelligence Laboratory has over 40 people working in various areas of artificial intelligence (machine learning, data mining, social network analysis, semantic technologies, computational linguistics, logic).

**Spinoff-s:** Quintelligence, Cyc-Europe, LiveNetLife, ModroOko, Envigence

**Academic Partners:** Carnegie Mellon, Cornell, Stanford, MIT, Uni. Maryland, KIT, UCL, ...

Jožef Stefan Institute, Artificial Intelligence Laboratory

Jozef Stefan Institute (JSI) is the leading Slovene research institution for natural sciences (900+ people) in the areas of computer science, physics, chemistry, ecology. Artificial Intelligence Laboratory has over 40 people working in various areas of artificial intelligence (machine learning, data mining, social network analysis, semantic technologies, computational linguistics, logic).

Spinoffs: Quintelligence, Cyc-Europe, LiveNetLife, ModroOko, Envigence

Academic Partners: Carnegie Mellon, Cornell, Stanford, MIT, Uni. Maryland, KIT, UCL, ...


Selection of Portals and Products:
- Text-Garden (http://www.textmining.net)
- Enrycher (http://enrycher.ijs.si/)
- VideoLectures.NET (http://videolectures.net/)
- IST-World (http://www.ist-world.org/)
- Search-Point (http://searchpoint.ijs.si/)
- OntoGen (http://ontogen.ijs.si/)
- Document-Atlas (http://docatlas.ijs.si/)
- Contextify (http://contextify.net/)
- NewsFeed (http://newsfeed.ijs.si/)
- DiversiNews (http://aidemo.ijs.si/diversinews/)
- EventRegistry (http://eventregistry.org/)
- Twitter Observatory (http://twitterobservatory.net/)

Selection of Projects (Integrated Projects and Networks of Excellence only):

Coordinating: XLike Cross-lingual Knowledge Extraction; Toposys Topological Complex Systems; NRG4Cast Energy Forecasting

H2020: MSCA RENOIR Reverse EngiNeering of sOcial Information pRocessing, MSCA BigDataFinance, OPTIMUM Multi-source Big Data Fusion Driven Proactivity for Intelligent Mobility, AQUASMART Aquaculture Open Data Cloud Innovation, CSA EDSA European Data Science Academy

IP: ACTIVE, COIN, EURIDICE, NeOn, ECOLEAD, SEKT

NoE: PlanetData, PASCAL2, MetaNet, Multilingual Web, LT-Web
Personal Background

Research in artificial intelligence, machine learning since 1987
Personal Background

Research in artificial intelligence, machine learning since 1987
PhD on machine learning methods for large text data
Personal Background

Research in artificial intelligence, machine learning since 1987
PhD on machine learning methods for large text data
Advising students and teaching courses at several universities
Personal Background

Research in artificial intelligence, machine learning since 1987
PhD on machine learning methods for large text data
Advising students and teaching courses at several universities
Leading Artificial Intelligence Laboratory at J. Stefan Institute (40+ people)

Leadership in science should **support excellence:**

- **everyone growing professionally and personally** in consciousness, gaining experience
- **no attachment to research results** fulfilling our expectations, see beyond the immediate situation (e.g., negative result may be lead to a new discovery) – long term vision
- **be flexible, adjust to new situations** (new research findings, new funding schemas, new problems, new technology,…)
- **see opportunity in every situation** - learn from experience, keep up
- **be authentic**, know and use your strengths and weaknesses
Personal Background

Research in artificial intelligence, machine learning since 1987
PhD on machine learning methods for large text data
Advising students and teaching courses at several universities
Leading Artificial Intelligence Laboratory at J. Stefan Institute (40+ people)

Leadership in science should **support excellence**:
- **everyone growing professionally and personally** in consciousness, gaining experience
- **no attachment to research results** fulfilling our expectations, see beyond the immediate situation (e.g., negative result may be lead to a new discovery) – long term vision
- **be flexible, adjust to new situations** (new research findings, new funding schemas, new problems, new technology, …)
- **see opportunity in every situation** - learn from experience, keep up
- **be authentic**, know and use your strengths and weaknesses

“…you will only be appreciated if you **appreciate the good side of your students**. If you want to tell someone they are rotten, just appreciate their good side instead. They will fully realize how rotten their other side is.” 

[Bhajan, 2001]
Science relies on people
Science relies on people

- **Science to be excellent** relies on excellent people having:
  - *Vision*, idea – have the goal bigger than you
  - *Courage*, enthusiasm for the vision – fight, cross the limits of known
Science relies on people

- **Science to be excellent** relies on excellent people having:
  - *Vision*, idea – have the goal bigger than you
  - *Courage*, enthusiasm for the vision – fight, cross the limits of known
  - *Grit*, persistency – stick with the goal and be ready to work hard
  - *Humility*, open mind – be prepared to listen to criticism/suggestions
Science relies on people

- **Science to be excellent** relies on excellent people having:
  - *Vision*, idea – have the goal bigger than you
  - *Courage*, enthusiasm for the vision – fight, cross the limits of known
  - *Grit*, persistency – stick with the goal and be ready to work hard
  - *Humility*, open mind – be prepared to listen to criticism/suggestions
  - *Knowledge* of the subject and related work – past and current study
  - *Prayer*, trust in the vision and the process – healthy openness for risk
Science relies on people

- **Science to be excellent** relies on excellent people having:
  - *Vision*, idea – have the goal bigger than you
  - *Courage*, enthusiasm for the vision – fight, cross the limits of known
  - *Grit*, persistency – stick with the goal and be ready to work hard
  - *Humility*, open mind – be prepared to listen to criticism/suggestions
  - *Knowledge* of the subject and related work – past and current study
  - *Prayer*, trust in the vision and the process – healthy openness for risk
  - *Grace* when handling difficulties – no losing time/energy fighting
  - *Determination*, strong decision – go for winning, put all of you in it, merge your vision/aim and your life

*Invest in knowledge, experience and developing talents of people.*
Graph/Social Network Analysis
(GraphGarden/SNAP, IST-World, FPIntelligence)

Data/Web/Text/Stream-Mining
(TextGarden Suite of tools)

Statistical Machine Learning
Computational Linguistics
(Enrycher, AnswerArt)

Complex Data Visualization
(DocAtlas, NewsExplorer, SearchPoint)

Graph/Social Network Analysis
(GraphGarden/SNAP, IST-World, FPIntelligence)

Data/Web/Text/Stream-Mining
(TextGarden Suite of tools)

Statistical Machine Learning
AILab Technologies

Social Computing/Web2.0 (LiveNetLife)

Computational Linguistics
(Enrycher, AnswerArt)

Complex Data Visualization
(DocAtlas, NewsExplorer, SearchPoint)

Graph/Social Network Analysis
(GraphGarden/SNAP, IST-World, FPIntelligence)

Data/Web/Text/Stream-Mining
(TextGarden Suite of tools)

Statistical Machine Learning
AILab Technologies

Deep Semantics & Reasoning (Cyc)

Light-Weight Semantic Technologies (OntoGen)

Social Computing/Web2.0 (LiveNetLife)

Computational Linguistics (Enrycher, AnswerArt)

Complex Data Visualization (DocAtlas, NewsExplorer, SearchPoint)

Graph/Social Network Analysis (GraphGarden/SNAP, IST-World, FPIntelligence)

Data/Web/Text/Stream-Mining (TextGarden Suite of tools)

Statistical Machine Learning
AlLab Technologies

Deep Semantics & Reasoning (Cyc)

Light-Weight Semantic Technologies (OntoGen)

Social Computing/Web2.0 (LiveNetLife)

Computational Linguistics (Enrycher, AnswerArt)

Complex Data Visualization (DocAtlas, NewsExplorer, SearchPoint)

Graph/Social Network Analysis (GraphGarden/SNAP, IST-World, FPIntelligence)

Data/Web/Text/Stream-Mining (TextGarden Suite of tools)

Statistical Machine Learning
Information Age - Age of Data Analytics

• Availability of large amounts of data → handling big data
  – millions of documents, sensor readings, astrophysics,…

• Data sources and variety of data → handling different data modalities
  – text understanding, genetics and molecular biology, video streams,…

• Data on different aspects of life → data science
  – fine-grained human behavior, interactions on social media,…

“This is the Information Age — everybody can be informed about anything and everything. There is no secret, therefore there is no sacredness. Life is going to become an open book. When your computer is more loyal, truthful, informed and excellent than you, you will be challenged. You do not have to compete with anybody. You have to compete with yourself.” [Bhajan, 2002]
Data Science

Interdisciplinary field, combines methods from
• statistics, machine learning, analytics,
• visualization,
• reporting, business intelligence, expert systems,
• databases, data mining, big data
Data Science

Interdisciplinary field, combines methods from

- statistics, machine learning, analytics,
- visualization,
- reporting, business intelligence, expert systems,
- databases, data mining, big data

Roles of people in Data science

- Project sponsor - business interest, championing the project
- Client - domain expert, end user
- Data scientist - set and execute analytics, managing the project
- Data architect - data management and storing
- Operations - acquiring data, infrastructure management, deployment
Big Data

From “Understanding Big Data” by IBM
Big Data

Data as asset on which to build business model

From “Understanding Big Data” by IBM
Big Data

Data as asset on which to build business model

Characterization of big data

From “Understanding Big Data” by IBM
Big Data

Data as asset on which to build business model

Characterization of big data

- **Volume** – data generated by machines, networks, social media, .... challenging to load and process (how to index, retrieve?)

From “Understanding Big Data” by IBM
Big Data

Data as asset on which to build business model

Characterization of big data

• **Volume** – data generated by machines, networks, social media, .... challenging to load and process (how to index, retrieve?)

• **Variety** – many sources and data types with different degree of structure (how to query semi-structured data?)

From “Understanding Big Data” by IBM
Big Data

Data as asset on which to build business model

Characterization of big data

- **Volume** – data generated by machines, networks, social media, .... challenging to load and process (how to index, retrieve?)
- **Variety** – many sources and data types with different degree of structure (how to query semi-structured data?)
- **Velocity** – continuous flow of data requires real-time processing influenced by rate of data arrival (how to handle high rate?)

From “Understanding Big Data” by IBM
Data Modalities
Data Modalities

- Cross-lingual text collections and text stream
  - Annotation using ontologies including word disambiguation
  - Transforming to triplets/RDF, to event templates, to logical form
  - Modeling evaluation of textual content in social networks
  - Knowledge acquisition for ontology extension
Text Stream

- News stream
- Documents
- User
- Ontology

1. Text enrichment using NLP & ML
2. Event template extraction
3. Relation extraction
4. User dialog for knowledge acquisition
5. Ontology extension
6. Text annotation
Data Modalities

- Cross-lingual text collections and text stream
  - Annotation using ontologies including word disambiguation
  - Transforming to triplets/RDF, to event templates, to logical form
  - Modeling evaluation of textual content in social networks
  - Knowledge acquisition for ontology extension
- Social media analysis
  - Modeling the user authority, modeling sentiment
- Structured data analysis for optimization
  - In combination with sensor data, with social media
Text and Social Networks

Social network

Evolution modeling (content, structure)

→ visualization
→ network properties
→ predictions
→ information diffusion

documents
Data Modalities

• Cross-lingual text collections and text stream
  – Annotation using ontologies including word disambiguation
  – Transforming to triplets/RDF, to event templates, to logical form
  – Modeling evaluation of textual content in social networks
  – Knowledge acquisition for ontology extension

• Social media analysis
  – Modeling the user authority, modeling sentiment

• Structured data analysis for optimization
  – In combination with sensor data, with social media

• Sensor data analysis
  – Visualization, prediction, semi-automatic pattern identification
  – In combination with social media and structured data for complex event detection
Sensors and Social Media Text

Social media

Complex event detection and prediction

Sensor network

Static data

event patterns

triggering alarms

ontology extension
Text Data Representation

- Different research fields work with textual data solving different problems:
  - Computational Linguistics, Machine Translation, Information Retrieval, Text Mining, Semantic Web, …

- Each of the research fields “represents” text in a slightly different way
How do we represent text?

- Character (character n-grams and sequences)
- Words (stop-words, stem/lemma)
- Phrases (word n-grams/skip-grams)
- Part-of-speech tags (noun, verb, adverb, …)
- Taxonomies / thesauri (WordNet)
- Vector-space model
- Correlated vector spaces (cross-lingual, cross-modal)
- Language models (probability of a word)
- Full-parsing (sentence parse tree)
- Collaborative tagging / Web2.0
- Templates / Frames
- Ontologies / First order theories
How do we represent text?

- Character (character n-grams and sequences)
- Words (stop-words, stem/lemma)
- Phrases (word n-grams/skip-grams)
- Part-of-speech tags (noun, verb, adverb,…)
- Taxonomies / thesauri (WordNet)
- Vector-space model
- Correlated vector spaces (cross-lingual, cross-modal)
- Language models (probability of a word)
- Full-parsing (sentence parse tree)
- Collaborative tagging / Web2.0
- Templates / Frames
- Ontologies / First order theories

Search, Categorization, Clustering, Summarization, …
How do we represent text?

- Character (character n-grams and sequences)
- Words (stop-words, stem/lemma)
- Phrases (word n-grams/skip-grams)
- Part-of-speech tags (noun, verb, adverb, ...)
- Taxonomies / thesauri (WordNet)
- Vector-space model
- Correlated vector spaces (cross-lingual, cross-modal)
- Language models (probability of a word)
- Full-parsing (sentence parse tree)
- Collaborative tagging / Web2.0
- Templates / Frames
- Ontologies / First order theories

Search, Categorization, Clustering, Summarization, …

Cross-lingual Inf. Retrieval, Connecting Text + Images,
How do we represent text?

- Character (character n-grams and sequences)
- Words (stop-words, stem/lemma)
- Phrases (word n-grams/skip-grams)
- Part-of-speech tags (noun, verb, adverb, …)
- Taxonomies / thesauri (WordNet)
- Vector-space model
- Correlated vector spaces (cross-lingual, cross-modal)
- Language models (probability of a word)
- Full-parsing (sentence parse tree)
- Collaborative tagging / Web2.0
- Templates / Frames
- Ontologies / First order theories

Search, Categorization, Clustering, Summarization, …

Cross-lingual Inf. Retrieval, Connecting Text + Images,

Machine translation, Spam filtering, …
Real-time Cross-lingual Global Media Monitoring

Real-time system based on ML and NLP enabling to:

- collect data from global media in real-time
- identify events and track evolving topics
- assign stable identifiers to events
- identify events across languages
- detect diversity of reporting along several dimensions
- provide rich exploratory visualizations
- provide interoperable data export

More in Leban, G., Fortuna, B., Brank, J., Grobelnik, M.,
Event Registry: Learning About World Events from News,
Global Media Monitoring pipeline

Input data

Pre-processing steps

- Article semantic annotation
- Extraction of date references
- Cross-lingual article matching
- Detection of article duplicates

Event construction

- Article clustering
- Cross-lingual cluster matching
- Event formation
- Event info. extraction
- Identifying related events

Event storage & maintenance

- Event registry

API Interface

GUI/Visualizations

Mainstream news

Blogs

http://EventRegistry.org

Input data

Pre-processing steps

- Article semantic annotation
- Extraction of date references
- Cross-lingual article matching
- Detection of article duplicates

Event construction

- Article clustering
- Cross-lingual cluster matching
- Event formation
- Event info. extraction
- Identifying related events

Event storage & maintenance

- Event registry

API Interface

GUI/Visualizations

Global Media Monitoring pipeline

Mainstream news

Blogs

http://EventRegistry.org
Global Media Monitoring pipeline

Input data
Mainstream news
Blogs

Pre-processing steps
Article semantic annotation
Extraction of date references
Cross-lingual article matching
Detection of article duplicates

Event construction
Article clustering
Cross-lingual cluster matching
Event formation
Event info. extraction
Identifying related events

Event storage & maintenance
Event registry
API Interface
GUI/Visualizations

Articles in different languages

http://EventRegistry.org
Global Media Monitoring pipeline

Input data

Pre-processing steps
- Article semantic annotation
- Extraction of date references
- Cross-lingual article matching
- Detection of article duplicates

Event construction
- Article clustering
- Cross-lingual cluster matching
- Event formation
- Event info. extraction
- Identifying related events

Event storage & maintenance
- Event registry

API Interface
- GUI/Visualizations

Main stream news

Blogs

Preprocessed and annotated enabling cross-lingual article matching

http://EventRegistry.org
Global Media Monitoring pipeline

Input data → Pre-processing steps → Event construction → Event storage & maintenance

Mainstream news:
- Article semantic annotation
- Extractions of date references
- Cross-lingual article matching
- Detection of article duplicates

Blogs:
- Article clustering
- Cross-lingual cluster matching
- Event formation
- Event info. extraction
- Identifying related events

Event registry → API Interface → GUI/Visualizations

Advanced search and rich visualization

http://EventRegistry.org
Related Systems/Demos

• NewsFeed (http://newsfeed.ijs.si/)
  – News and social media crawler
• Enrycher (http://enrycher.ijs.si/)
  – Language and Semantic annotation
• SearchPoint (http://searchpoint.ijs.si/)
  – Contextualized search
• XLing (http://xling.ijs.si/)
  – Cross-lingual document linking and categorization
• Event Registry (http://eventregistry.org/)
  – Event detection and topic tracking
Collecting global media data

- Data collection service News-Feed
  - [http://newsfeed.ijs.si/](http://newsfeed.ijs.si/)
  - …crawling global main-stream and social media

- Monitoring
  - ~70k main-stream publishers (RSS feeds + special feeds)
  - ~250k most influential blogs (RSS feeds)
  - free Twitter feed

- Data volume: ~350k articles & blogs per day (+5M tweets)
- Languages: eng (50%), ger (10%), spa (8%), fra (5%),…
How can we annotate a document?

<table>
<thead>
<tr>
<th>Level</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical level</td>
<td><strong>Tokenization</strong> – extracting tokens from a document (words, separators, …)</td>
</tr>
<tr>
<td></td>
<td><strong>Sentence splitting</strong> – set of sentences to be further processed</td>
</tr>
<tr>
<td>Linguistic level</td>
<td><strong>Part-of-Speech</strong> – assigning word types (nouns, verbs, adjectives, …)</td>
</tr>
<tr>
<td></td>
<td><strong>Deep Parsing</strong> – constructing parse trees from sentences</td>
</tr>
<tr>
<td></td>
<td><strong>Triple extraction</strong> – subject-predicate-object triple extraction</td>
</tr>
<tr>
<td></td>
<td><strong>Name entity extraction</strong> – identifying names of people, places, organizations</td>
</tr>
<tr>
<td>Semantic level</td>
<td><strong>Co-reference resolution</strong> – replacing pronouns with corresponding names; merging different surface forms of names into single entity</td>
</tr>
<tr>
<td></td>
<td><strong>Semantic labeling</strong> – assigning semantic identifiers to names (e.g. LOD/DBpedia/Freebase) including disambiguation</td>
</tr>
<tr>
<td></td>
<td><strong>Topic classification</strong> – assigning topic categories to a document (e.g. DMoz)</td>
</tr>
<tr>
<td></td>
<td><strong>Summarization</strong> – assigning importance to parts of a document</td>
</tr>
<tr>
<td></td>
<td><strong>Fact extraction</strong> – extracting relevant facts from a document</td>
</tr>
</tbody>
</table>
Slovenia’s dramatic win over Russia Wednesday, and to a lesser extent Ireland’s narrow loss to France, capped off a grueling two-year qualifying period that saw some of the smallest countries in the world kick some of soccer’s biggest names in the teeth. After a century of notching victories from the likes of Brazil, Italy and Germany, international soccer is entering the era of the Cinderella. It may not happen this way, but given the increasing flow of talent, training and information across borders, it’s almost certain that a small upstart nation blessed with good athletes and better luck will make a legitimate run at the world’s most coveted trophy.

Russia’s Yuri Zhirkov, right, fights for the ball with Slovenia’s Valter Birsa Wednesday.

“Enrycher” is available as a web-service generating Semantic Graph, LOD links, Entities, Keywords, Categories, Text Summarization, Sentiment

Diego Maradona Semantics:
owl:sameAs: http://dbpedia.org/resource/Diego_Maradona
owl:sameAs: http://sw.opencyc.org/concept/Mx4rvofERZwpEbGdrcN5Y29ycA
rdf:type: http://dbpedia.org/class/yago/ArgentinaInternationalFootballers
rdf:type: http://dbpedia.org/class/yago/ArgentineExpatriatesInItaly
rdf:type: http://dbpedia.org/class/yago/ArgentineFootballManagers
rdf:type: http://dbpedia.org/class/yago/ArgentineFootballers

Robbie Keane Semantics:
owl:sameAs: http://dbpedia.org/resource/Robbie_Keane
rdf:type: http://dbpedia.org/class/yago/CoventryCityF.C.Players
rdf:type: http://dbpedia.org/class/yago/ExpatriateFootballPlayersInItaly
rdf:type: http://dbpedia.org/class/yago/F.C.InternazionaleMilanoPlayers
**Enrycher** is a web service consisting of a set of interlinked modules

- covering lexical, linguistic and semantic annotations
- exporting data in XML or RDF

To execute the service, one should send an HTTP POST request, with the raw text in the body:

```bash
- curl -d "Enrycher was developed at JSI, a research institute in Ljubljana. Ljubljana is the capital of Slovenia."
  http://enrycher.ijs.si/run
```
SearchPoint - Contextualized search
(http://searchpoint.ijs.si/)

“Contract”

Graphical User Interface

SABLE

METADATA

DOCUMENTS

Context Discovery
Soft Filtering
Graph Drawing
SearchPoint - Contextualized search (http://searchpoint.ijs.si/)

Accenture - Sable SearchPoint

SABLE

“Contract”

METADATA

DOCUMENTS

Context Discovery

Soft Filtering

Graph Drawing
Text Representation for Cross-lingual Data Analytics

• Represent text in a language-neutral form based on statistical methods
  – document content is comparable regardless of the natural language of the documents

• Useful for different problems involving information retrieval, classification, clustering, …

• We can solve this on a large scale
  – also because of availability of large amounts of “comparable corpora” like Wikipedia or Acquis (EU legislation)
Correlation analysis

Language Neutral Document Representation (trained with machine learning)
Correlation analysis

Language Neutral Document Representation (trained with machine learning)

New document represented as text in any of the above languages
Language Neutral Document Representation (trained with machine learning)

New document represented as text in any of the above languages
Correlation analysis

Language Neutral Document Representation (trained with machine learning)

New document represented as text in any of the above languages

New document represented in Language Neutral way
New document represented as text in **any** of the above languages

New document represented in **Language Neutral** way

...enables cross-lingual retrieval, categorization, clustering, ...
With machine learning techniques we can learn “language neutral document representation”…

…for over 100 Wikipedia languages each having over 10 000 articles

Wikipedia Languages

• With machine learning techniques we can learn “language neutral document representation”…

• …for over 100 Wikipedia languages each having over 10,000 articles

More in A. Muhić, J. Rupnik, P. Škraba.
Cross-Lingual Document Retrieval through Hub Languages, xLiTe: Cross-Lingual Technologies, NIPS 2012 Workshop.
Wikipedia Languages

- With machine learning techniques we can learn "language neutral document representation"…

- …for over 100 Wikipedia languages each having over 10 000 articles

Document representation

Write each document in aligned Wikipedia basis (index documents)
Event Registry system for global media monitoring ([http://eventregistry.org](http://eventregistry.org))

- Having a stream of news & social media, the task is to structure documents into events
- Event Registry allows for:
  - Identification of events from documents
  - Connecting documents across many languages
  - Tracking events and constructing story-lines
  - Describing events in a (semi)structured way
  - UI for exploration through Search & Visualization
  - Export into JSON/RDF (Storyline ontology)
“Event Registry” example on “Chicago” related events (http://eventregistry.org)
Migrants Plot Alternative Routes as Hungary Detains 500

WHEN: September 20, 2015
WHERE: Zagreb, Croatia

LONDON -- Hundreds of migrants remained stranded on Serbia's border with Hungary early Wednesday as Hungary's decision to seal its border rippled across Europe and other migrants scrambled to find alternative routes, in an effort, in most cases, to reach Germany. Some were planning to go through Croatia and Slovenia. Another possible route is through Hungary's border with Romania -- which, however, Hungary is also moving to tighten -- or the land border between Turkey and Bulgaria. At a bus...

Pope Francis and city of Washington embrace each other on historic visit

WHEN: September 23, 2015
WHERE: New York, United States
Top concepts in events

List of concepts (entities and keywords) that best describe what the events are about

<table>
<thead>
<tr>
<th>ENTITIES</th>
<th>Concept relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syria</td>
<td>21</td>
</tr>
<tr>
<td>Germany</td>
<td>17</td>
</tr>
<tr>
<td>Europe</td>
<td>15</td>
</tr>
<tr>
<td>European Union</td>
<td>14</td>
</tr>
<tr>
<td>United States</td>
<td>14</td>
</tr>
<tr>
<td>United Nations</td>
<td>13</td>
</tr>
<tr>
<td>Turkey</td>
<td>10</td>
</tr>
<tr>
<td>Iraq</td>
<td>9</td>
</tr>
<tr>
<td>Italy</td>
<td>7</td>
</tr>
<tr>
<td>United Nations High Commissioner</td>
<td>6</td>
</tr>
<tr>
<td>Hungary</td>
<td>5</td>
</tr>
<tr>
<td>Lebanon</td>
<td>5</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5</td>
</tr>
<tr>
<td>Greece</td>
<td>5</td>
</tr>
<tr>
<td>France</td>
<td>5</td>
</tr>
<tr>
<td>Berlin</td>
<td>5</td>
</tr>
<tr>
<td>Austria</td>
<td>4</td>
</tr>
<tr>
<td>Social Democratic Party of Germany</td>
<td>4</td>
</tr>
<tr>
<td>Christian Democratic Union...</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KEYWORDS</th>
<th>Concept relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refugee</td>
<td>45</td>
</tr>
<tr>
<td>Asylum seeker</td>
<td>11</td>
</tr>
<tr>
<td>Immigration</td>
<td>11</td>
</tr>
<tr>
<td>Right of asylum</td>
<td>9</td>
</tr>
<tr>
<td>States of Germany</td>
<td>8</td>
</tr>
<tr>
<td>Government</td>
<td>7</td>
</tr>
<tr>
<td>Police</td>
<td>7</td>
</tr>
<tr>
<td>Mediterranean Sea</td>
<td>5</td>
</tr>
<tr>
<td>Euro</td>
<td>5</td>
</tr>
<tr>
<td>Politics</td>
<td>5</td>
</tr>
<tr>
<td>Terrorism</td>
<td>4</td>
</tr>
<tr>
<td>Human rights</td>
<td>4</td>
</tr>
<tr>
<td>Islam</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>4</td>
</tr>
<tr>
<td>City</td>
<td>4</td>
</tr>
<tr>
<td>Law</td>
<td>4</td>
</tr>
<tr>
<td>Violence</td>
<td>3</td>
</tr>
<tr>
<td>Middle East</td>
<td>3</td>
</tr>
<tr>
<td>Border</td>
<td>3</td>
</tr>
<tr>
<td>Gulf</td>
<td>2</td>
</tr>
</tbody>
</table>

WHERE: New York, United States
Clustering of events

The dendrogram displays how events can be split into subgroups based on their similarity. Clicking the blue dots will display sub-clusters. Clicking buttons with the number of events will display the events in the cluster.
38.9% events belong to this category.
The Artificial Intelligence Laboratory is concerned mainly with research and development in information technologies with an emphasis on artificial intelligence. The main research areas are the following: (a) data analysis with an emphasis on text, web and cross-modal data, (b) scalable real-time data analysis, (c) visualization of complex data, (d) semantic technologies, (e) language technologies.

In collaboration with the Department of Communication Systems (E6) and Centre for Knowledge Transfer in Information Technologies (CT3) we have established a Cross-department laboratory for wireless sensor networks (SensorLab). The goal is to combine technologies for (a) sensor data acquisition, (b) communication between sensor devices, (c) statistical real-time data analysis, (d) semantic technologies, and to enable a wide range of research and development in different application areas, such as energy, ecology, transport, security, and logistics.

The Artificial Intelligence Laboratory puts special emphasis on the promotion of science. In collaboration with the Centre for Knowledge Transfer in Information Technologies (CT3) we are developing the VideoLectures.NET educational portal and organizing the national ACM competition in Computer Science (in Slovene).

The Artificial Intelligence Laboratory has a well-established collaboration with a number of academic and commercial organizations, some members of the Laboratory are involved with Stanford University, University College London, Jožef Stefan International Postgraduate School and companies Quintelligence, Cycorp Europe, LifeNetLive, Modro Oko and Envigence.
The Artificial Intelligence Laboratory is concerned mainly with research and development in information technologies with an emphasis on artificial intelligence. The main research areas are the following: (a) data analysis with an emphasis on text, web and cross-modal data, (b) scalable real-time data analysis, (c) visualization of complex data, (d) semantic technologies, (e) language technologies.