PANEL IMMM/ICIW

Bridging the Web of Data with the Web of Documents
Information

Information: discovery, interpretation
(semantic, ontology, formal/non-formal)

Nature: text, photos, voice

Representation: text, tables, graphics, functions

Context
BIG/SMALL Data
Linked Data
Open Data
Panel

• **Moderator**
  > Petre Dini, Concordia University, Canada/ China Space Agency Center, China

• **Panelists**
  > Awatef Hicheur Cairns, Altran Research, France
    Is it more appropriate to use process discovery techniques or sequence mining techniques to extract important patterns in students’ behaviors and to make recommendations for the best learning paths for students to follow

  > António Teixeira, University of Aveiro, Portugal
    Applications of extraction of semantic information from publicly available documents regarding Health and local government

  > Athanassios Jimoyiannis, University of Peloponnese, Greece
    Participative-learning affordances of Social Media and Web 2.0, e.g. Participatory Web, Open Web, Collaborative Web, and Social Web.
    A new pedagogical and learning context, which combines formal, non-formal and informal features; e.g., autonomous and self-directed learning, collaborative learning and ubiquitous learning, networking and community learning.
The need and applications of extraction of semantic information from publicly available documents

(with concrete examples for Health and eGov)

António Teixeira
Dep Electronics Telecom. & Informatics/IEETA
University of Aveiro
Governments produce large amounts of documents

– At the several levels (local, central ...)
– Of many types:
  • Laws
  • Regulations
  • Minutes of meetings
  • Deliberations
  • ...
Problem(s)

- Relevant information produced in (written) natural language
- How to make available as really usefull information for citizens?
- How to make it usable by machines?
Face to face interaction:
- Working hours
- In loco
- Relevant Information

Keyword information search:
- Anytime
- Anywhere
- Difficult to find relevant data in large result set

Search that understands language:
- Anytime
- Anywhere
- Relevant Information
More and more healthcare institutions store vast amounts of information:
- about users, procedures, and examinations, as well as the findings, test results, and diagnoses respectively.

Other institutions, such as the Government, increasingly disclose health information on varied topics of concern to the public writ large.

Health research is one of the most active areas, resulting in a steady flow of publications reporting on new findings and results.

In recent years, the Internet has become one of the most important tools to obtain medical and health information.
Examples of data - Health

- Public
  - Medical journals
  - Newspapers
  - Masters and PhDs
  - Books
  - ...

- Not public
  - Hospital Recordings
  - Doctors documents
  - ...

IMMM/ICIW Panel - Bridging the Web of
General search engines do not allow the end-user to obtain a clear and organized presentation of the available health information.

- Instead, it is more or less of a hit or miss, random return of information on any given search.
- In fact, medicine-related information search is different from other information searches, since users often use medical terminology, disease knowledge, and treatment options in their search (Wang et al. 2012).

Much of the information that would be of interest to private citizens, researchers, and health professionals is found in unstructured documents.
The Gap between Data/Docs and Humans (and Machines)
The need for semantic search

- Efficient access to this information implies the development of search systems capable of handling the technical lexicon of the domain area, entities such as drugs and exams, and the domain structure.

- Such search systems are said to perform semantic search as they base the search on the concepts.

- Semantic search maintains several advantages over search based on surface methods:
  - such as those that directly index text words themselves rather than underlying concepts.

- Three main advantages of concept-based search are:
  - they usually produce smaller sets of results, as they are able to identify and remove semantically duplicated results and/or semantically irrelevant results;
  - they can integrate related information scattered across documents; frequently answers are obtained by compounding information from two or more sources;
  - they can retrieve relevant results even when the question and answer do not have common words, since these systems can be aware of similar concepts, synonyms, meronyms, antonyms, etc.
But the semantic information must be derived from the documents...

- Using techniques from Information Extraction, NLP...

Obtained information must be easy to explore

Complex and more natural queries must be possible
Examples

From my Recent Work
Chapter: Online Health Information Semantic Search and Exploration: Reporting on Two Prototypes for Performing Information Extraction on both a Hospital Intranet and on the World Wide Web

IMMM/ICIW Panel - Bridging the Web of Data with the Web of Documents
Ontology-based Health information Search (in project World Search)

System consists of three principal modules:
   a) semantic annotation;
   b) semantic navigation interface; and
   c) search platform
This tool allows navigation of the semantic information.

When users search for something, the platform analyzes the query and verifies in the knowledge base the semantic of this text, constructing a knowledge graph.
Example of Use

Memantina is one of the terms identified in the tag cloud selected.

If the user wants to know what memantina is or what its relationship is with these diseases and risk factors, selecting memantina the system will add this term to the query.
eGov Scenarios – Conceptual Model

Semantic Extraction & Integration

Natural Language Processing

responsible service

knowledge representation

user

Semantic Web

doc. A

doc. B

doc. C

... structured sources

www
Use scenarios
Thank you for your attention.

Acknowledgments

- Mário Rodrigues and Liliana Ferreira
- Part of the mentioned work was part of World Search project, a QREN project (QREN 11495), co-funded by COMPETE and FEDER.
- Also acknowledged the support by IEETA Research Unit, FCOMP-01-0124-FEDER-022682 (FCT-Pest C/EEI/UI0127/2011).
Data mining and Process Mining in the educational field

Awatef Hicheur Cairns
Altran Reaseach
Contact : awatef.hicheurcairns@altran.com
Educational data Mining

- Specific data mining methods to
  - explore the unique types of educational datasets
  - get a better understanding of students and the settings on which they learn

- Educational data mining methods
  - Prediction
  - Clustering
  - Relationship Mining
  - Discovery with Models
  - Distillation of Data For Human Judgment

- Limitation of the data mining techniques
  - Data centric, not process centric
  - Focus on data or sequential structures rather than whole process models with concurrency patterns
Positioning Process Mining

process model analysis
(simulation, verification, etc.)

process mining

data-oriented analysis
(data mining, machine learning, business intelligence)

performance-oriented questions, problems and solutions

compliance-oriented questions, problems and solutions
Process Mining spectrum

1) process model

2) organizational model

3) social network

Model extraction

4) basic performance metrics

5) operational support

6) auditing/security

7) model extension

Extending models with frequencies, temporal information, decision points

4) basic performance metrics

Conformance checking

If …then …

Conforming predictive models

Operational support (prediction, recommendation, etc.)

Replay model against event log

Constructing predictive models
Process mining: Operational support

- Possible recommendations:
  - next activity;
  - suitable resource; or
  - routing decision.

- A recommendation is always given with respect to a specific goal. Examples of goals are:
  - minimize the remaining flow time;
  - minimize the total costs;
  - maximize the fraction of cases handled within 4 weeks;
  - maximize the fraction of cases that is accepted; and
  - minimize resource usage.
Process Mining challenges in the educational field

✓ **Voluminous Data** - Large number of cases or events in event logs
  ➢ Possible solution: clustering techniques to partition logs and distributed algorithms to parallelize computation

✓ **Heterogeneity and Complexity** - Large number of distinct traces and activities in event logs
  ➢ Possible solution: filtering, abstraction or clustering techniques may help reducing the complexity

✓ **Concept drift** - Educational processes may change while being analyzed
  ➢ Possible solution: splitting the event log into smaller logs

✓ **Usability and understandability** for end users
  ➢ Proposed solutions: visualization techniques and notation simplification
Thank you for your attention!

Questions?
The three main types of process mining: discovery, conformance, and enhancement
Exemple of process discovery algorithms

- **Algorithmic techniques**
  - Alpha miner
  - Alpha+, Alpha++, Alpha#
  - FSM miner
  - Fuzzy miner
  - Heuristic miner
  - Multi phase miner

- **Genetic process mining**
  - Single/duplicate tasks
  - Distributed GM

- **Region-based process mining**
  - State-based regions
  - Language based regions

- **Classical approaches not dealing with concurrency**
  - Inductive inference (Mark Gold, Dana Angluin et al.)
  - Sequence mining
Process discovery Challenge: Four Competing Quality Criteria

“able to replay event log”

fitness

generalization

“Occam’s razor”

simplicity

precision

process discovery
Decision tree learning
Clustering
Association rule learning

Règle de la forme “IF X THEN Y” \( X \Rightarrow Y \).

<table>
<thead>
<tr>
<th></th>
<th>cappuccino</th>
<th>latte</th>
<th>espresso</th>
<th>americano</th>
<th>ristretto</th>
<th>tea</th>
<th>muffin</th>
<th>bagel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\( tea \land latte \Rightarrow muffin \)

\( tea \Rightarrow muffin \land bagel \)
Sequence mining

$$X \implies Y$$

<table>
<thead>
<tr>
<th>customer</th>
<th>seq. number</th>
<th>timestamp</th>
<th>items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wil</td>
<td>1</td>
<td>02-01-2011:09.02</td>
<td>{cappuccino}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>03-01-2011:10.06</td>
<td>{espresso,muffin}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>05-01-2011:15.12</td>
<td>{americano,cappuccino}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>06-01-2011:11.18</td>
<td>{espresso,muffin}</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>07-01-2011:14.24</td>
<td>{cappuccino}</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>07-01-2011:14.24</td>
<td>{americano,cappuccino}</td>
</tr>
<tr>
<td>Mary</td>
<td>1</td>
<td>30-12-2010:11.32</td>
<td>{tea}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30-12-2010:12.12</td>
<td>{cappuccino}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>30-12-2010:14.16</td>
<td>{espresso,muffin}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>05-01-2011:11.22</td>
<td>{bagel,tea}</td>
</tr>
<tr>
<td>Bill</td>
<td>1</td>
<td>30-12-2010:14.32</td>
<td>{cappuccino}</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30-12-2010:15.06</td>
<td>{cappuccino}</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>30-12-2010:16.34</td>
<td>{bagel,espresso,muffin}</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>06-01-2011:09.18</td>
<td>{ristretto}</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>06-01-2011:12.18</td>
<td>{cappuccino}</td>
</tr>
</tbody>
</table>

$$X = \langle \{cappuccino\}, \{espresso\} \rangle$$
$$Y = \langle \{cappuccino\}, \{espresso\}, \{latte,muffin\} \rangle$$
Episode mining

32 time windows of length 5
Towards the Web of Learning

Panel “Bridging the Web of Data with the Web of Documents”

Prof. Athanassios Jimoyiannis
Department of Social and Educational Policy
University of Peloponnese, Greece

Paris, France
The Web of documents

Institution Website

Department Website

Department LMS

Class blog

Class wiki

Course site

Course and class forums

Instructor's blog

Students’ personal blogs

Personal Twitter Facebook...
Web as a learning platform: e-Learning 2.0

- **Openness**
  - Open Educational Resources
  - User generated content

- **Participation**
  - Emergent, self-directed learning

- **Interactivity**
  - Sharing ideas
  - Discussion
  - Reflection

- **Sociability**
  - Learning beyond the classroom boundaries
  - Communities of learning

- **Semantic**
  - Content aggregation and reusability

- **Collaboration and creativity**
  - Content sharing and transforming
  - Co-creating new concepts & forms
The blended structure of Web-based learning

- **Formal learning**
  - Educational settings
    - Lecture, presentation FtF activities
  - Professional content
- **Community learning**
  - On-line spaces
    - Communication
    - Reflection, content sharing and co-creation
  - Student professional content
- **Informal learning**
  - Personal space
    - Engagement
    - Self-directed activities
  - Student performance content
From data to competence

Social dimension

Data

Situated, social learning
Sharing, co-creation
User generated
Reflecting
Communicating
Gathering

Content

Understanding
Comprehension
Integration
Application

Knowledge

Competition
abilities, innovation, attitudes, values

Personal dimension

User generated

Situated, social learning
Sharing, co-creation
User generated
Reflecting
Communicating
Gathering

Understanding
Comprehension
Integration
Application

innovation
Towards Pedagogy 2.0

- Beyond didactic paradigm (knowledge transfer)
- Beyond constructivism
- Adding Community to the Content
- Shifting the focus from the teaching-content of a subject to the learning activities
  - Shared content and resources
  - Collaborative learning
  - Learners as content creators
  - Networked and collective intelligence
  - Self-directed and ubiquitous learning
Learning 2.0: Open questions

- How do we design educational programs and learning tasks through Web-based, social learning approaches?
- What are the best practices to support and scaffold students in a system where learning is expected outside of school settings?
- Are instructors ready - properly prepared to effectively respond to their changing role?
“No teacher left behind!”

Thanks to the participants who contributed to this topic/debate with their criticism, comments, ideas and experiences.