Relevance of the topic for Linguistists

- A new area is emerging which requires linguistic expertise
  - Socio-semantic computing is a new prospective area for application of your expertise
    - It is mature, it is an area of growing commercial interest
    - but new for linguists

- Methods, I’ll talk about, are useful for exploiting ontologies and for natural language understanding
  - Everything is deeply intertwined. In an important sense there are no "subjects" at all; there is only all knowledge, since the cross-connections among the myriad topics of this world simply cannot be divided up neatly.
    Theodor Holm Nelson, the pioneer of information technology.
  - and nowadays many of this links are explicit in computer mediated communication networks (the web, semantic web, desktop data, intranets, ...), which can be studied using interdisciplinary methods
  - We agree with J.R. Firth "You shall know the word by the company it keeps", then nowadays we can know the text by the company it keeps in the unified socio-semantic space
Social Semantic Web

- **Semantic Web**
  - Focuses on data interoperability across applications and organisations
  - by means of top-down constructed formal ontologies

- **The proliferation of Web 2.0**
  - Brought about all kinds of digital artefacts: documents, people, concepts, vocabulary, tasks, activities, and more. The Web is increasingly becoming a participatory, social space
  - Established tagging as popular mechanism to replace hierarchical categorization and formal ontologies

- **Socio-Semantic Web (Web 3.0?) is emerging:**
  - Focuses on personalization, small and massive scale collaboration, findability and navigation
  - Using collaboration environments that exploit the semantics of the open content (ranging from Wikipedia articles to tags)
    - Text understanding and bottom-up conceptualization
Social Semantic Web

• **Challenges:**
  • How can computers help with realizing massive collaboration?
  • How to filter, aggregate, align, weight, and incorporate millions of single contributions in a sensible way?
  • Can we exploit the semantics of a contribution in order to integrate it in the previous contributions?
  • ...

• **I’ll talk about some problems and solutions for Socio-Semantic Web**
  • I’ll illustrate these as the use cases of Galaxy library recently created in IBM: IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks
    • Provides a unified API that helps in creating solutions for these types of multidimensional networks (people, documents, tasks, etc.)
    • Provides an integrated platform for combining social computing, semantic processing, and activity-centered computing for enhanced user experience.
IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks

The proliferation of Web 2.0 has brought about all kinds of digital artifacts: documents, people, concepts, vocabulary, tasks, activities, and more. IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks provides a unified API that helps in creating solutions for these types of multidimensional networks (people, documents, tasks, etc.) and provides an integrated platform for combining social computing, semantic processing, and activity-centered computing for enhanced user experience.

http://www.alphaworks.ibm.com/tech/galaxy
Use Case:

Navigation which is ambient
John is looking for collaboration

John B.

John is looking for a mentor
Galaxy gets input:
Galaxy locates object around John and suggests to connect with Tim

John B.  

?  

Tim B.
Tim is three steps away from John
“Three steps away” 😞?

Why Galaxy decided that this three steps away connection is a strong connection?
John and Tim –

Galaxy computes that this is a strong connection because of multiple ways of connections.
John and Derek

Galaxy computes that such type of connectivity is a weak connection
Use Case:
Ambient Navigation
in Social Semantic Desktop
(Created in Nepomuk)
Dirk has been given a task
Galaxy provides a solution

Dirk

Project

Documentation

Ansgar

Report
Dirk wonders “Why me?”
Right person

Galaxy computes that this is a strong connection because of multiple ways of connections
Use Case:

Finding new facts for an ontology

using that ontology
Finding indirect links that could be relevant

- In the scenario “John is looking for a mentor” we actually demonstrated that through mining a graph representing relations between people and the things they create and do, Galaxy found a new useful fact:
  - [John] is related to [Tim]
    by the relation “strong multiple connectivity”

- Superficially, Galaxy UI performs like Google Sets:
  - User focuses on one or more concepts, like [John], and Galaxy tries to predict other concepts of interest.
Advantages of using Galaxy

• No SPARQL or other queries
• No browsing
• Fast
• Shows something of cognitive interest
  • to perceive, contextualize, simplify, and make sense of otherwise complex interlinked data
  • without cognitive load: How can I ask for what?
Using Galaxy for ego-centric queries

• On the next slides we show how Galaxy performs for egocentric queries
  • finds another instances of the same class, provides generalisation, finds common features of “neighbours”
  • and ranks these findings in a sensible way, depending on the topology of the cognitive map
  
• With whom is Claudia connected?
  • With Dirk, Martin, Elain, John, Hanna, etc?
  • With “some researchers”, like Dirk, Martin, Elain, John, Hanna, etc?
  • With “many researchers”? 
UI of Galaxy library
Generalisation

... → Cats → Sappho, Mischa, Minka → ...

...
With whom is Claudia connected?

All of these people

Claudia

Dirk

Martin

Elaine

John

Hanna

Researcher
Ranking
Ranking

1

2

...
Summary:

What do we mean by

“Ambient Navigation”
“Ambient Navigation” is our user-centric generalisation of “dynamic taxonomies”

- multidimensional networks (like PIMO) provide a single, coherent framework in which users can focus on one or more nodes (concepts) in the network, and immediately see a conceptual summary of their focus,
  - in the form of a reduced network derived from the original one by pruning unrelated concepts
  - augmented with relations “strong connectivity”

- Concepts in the transformed network can be used to set additional, dependent foci and users iterate in a guided yet unconstrained way until they reach a result set sufficiently small for manual inspection
Use Case:

Collaborative Tagging Systems

(dogear, Del.icio.us, …)
Data is a (multidimensional) network
Galaxy provides:

- Community detection
- Community-based tag recommendation
- Expertise location
- ...

Why Galaxy might be a better solution than alternatives?

- Fast
- Takes into account multiple relations (as illustrated on the next slide)
… taking into account multiple relations
Problems of large scale collaborative tagging

- **Flat “tag soup” of ambiguous tags**
  - Prevents scalability and
  - damages findability
- **Approaches**
  - how this problem will be addressed by researchers and industry?
  - By introducing some structures for tags:
    - bundles, classifications, relations, tagging of tags, ...
  - and by disambiguation
- **We propose to use Galaxy to**
  - cluster tags, and possible tag’s meanings
  - disambiguate ambiguous tags
  - For example, using Galaxy we find that there are two strong clusters in the space of people and resources related to tag **BP**
    - **British Petroleum**
    - **IBM Blue pages**
Use Case:

Analysis of Massive Social Networks
Data Integration
This is a small Social Network
a bigger Social Network
Very big Social Network

where individual actors (people) are not visible
Galaxy: numerical simulation

Galaxy can detect sub-communities and central people (red dots) in these communities on different levels of granularity: mega, mezzo, micro.
Galaxy: massive SN

and provide structured view of massive networks (for example, as topic maps)
Visualisation as the use case:

Results of this analysis allows to provide visualisation which is both “beautiful” + “structured”:
- Sub-communities are visible
- Central people are highlighted

Visualisation becomes fast, because Galaxy can be used to create a reduced network by pruning concepts, which are too far from the user focus (foci)
Introducing Temporal Aspects into Modelling of Massive Dynamic Networks
Mining dynamic networks

- **Time affects the needs of socio-semantic network applications**
  - Ad-hoc communities of interest are formed and dissolved
  - Events are tied to particular points in time when they occur
  - Metadata (like tags in Del.icio.us) might “expire”, and not fit to current cognitive needs of the user

- **Numerous collections of longitudinal data, or data with timestamps (phone calls registered by telephone companies, emails) “beg for mining” which will allow**
  - to identify trends
  - to predict future
  - to find recurring patterns of events, etc
  - to identify social roles of actors

- **In order to**
  - set competitive tariffs for traffic,
  - create better tools for workflow management
  - provide time-sensitive search and navigation, etc
IBM/Nepomuk Galaxy library aims to provide a unified API that helps in creating solutions for socio-semantic applications.

We investigate how to use Galaxy for modeling of temporal aspects, including considerations on:

- Data representation
- Graphmining techniques
- Goals of mining and potential range of application’s domains

We are now exploring two quite distinctive approaches, which are distinctive based on these criteria:

- synchronic study,
- as opposed to examining the temporal development of the network.
Two approaches to model the temporal dimension

- **Synchronic**
  - Study the network at particular time slices
  - Time points are nodes linked to each other and to events etc
  - Straight forward
  - Simple data aggregation

- **Diachronic**
  - Network and its linkage is dynamic
  - Strength of certain inter node links decay or increase over time e.g. as ad-hoc communities are formed or dissolve
  - More complicated
  - Data is changed rather than aggregated
Synchronic approach allows to study longitudinal data
Synchronic approach allows to study longitudinal data
Synchronic approach allows to study longitudinal data
Synchronic analysis & applications

• **Detect structure of networks at different time slices**
  - networks with complex topology (not necessarily “grids” as in image processing)

• **Applications:**
  - trend analysis, risk assessment
    - The component “left leg” became more prominent
    - The component “Head” became less prominent
Diachronic analysis & applications

• **Diachronic**
  - Events are tied to particular points in time (timestamps) when they occur
  - Graph mining algorithms can move through time
  - Recurring patterns can be anticipated by studying events linked to the same or close points in time

• **Applications:**
  - Activity centric computing, workflow management
Use Case:

Activities Detection
Diachronic – Event Anticipation

Event: Calendar is opened

Event: Workshop is discussed

Event: Travel Reservation is accessed

Probably, user needs the link to Travel Reservation
How to take your next steps

• Read Galaxy “Getting started”
  • Download from: http://www.alphaworks.ibm.com/tech/galaxy

• Learn Galaxy’s news:
  • http://atroussov.com/Activities.html

• Contact me:
  • atrousso@ie.ibm.com
  for research collaboration and for job opportunities in IBM
Acknowledgements

• Acknowledgements:
  • my IBM collaborators: Mikhail Sogrin and John Judge
  • Nepomuk project partners:
    • Brian Davis and Siegfried Handschuh (NUIG)
    • Alexander Polonsky and Mikhail Kotelnikov (Cognium Systems)
    • Ying Du, Ernie Ong, and Uwe Riss (SAP Research Team)
    • Leo Sauerman and Gunnar Grimnes (DFKI),
    • Heiko Haller and Max Voelkel (FZI)