Mining Multidimensional Socio-Semantic Networks

03.03.2008
Guest lecture for students of Network and Distributed System MSc class in Trinity College Dublin

Alexander Trousov, Ph.D.
IBM Dublin Center for Advanced Studies
Chief Scientist

http://nepomuk.semanticdesktop.org
Agenda

- Part 1 Social Semantic Web and the Galaxy Library
  - Introduction
  - Use Cases
    - Ambient Navigation
    - Ambient Navigation (Nepomuk)
    - Ontology Mining
    - Collaborative/Social Tagging
    - Analysis of Massive Social Networks
Agenda

- Part 2 Modelling Temporal Aspects
  - Time as a dimension
  - Synchronic and Diachronic
  - Mining Temporal Data

This slide is a draft
(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**
  - has brought about all kinds of digital artifacts: 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 http://www.alphaworks.ibm.com/tech/galaxy

  - which 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.
Use Case:

Navigation which is ambient
John is looking for collaboration

John B.

John is looking for a mentor
Galaxy suggests to connect with Tim

John B.  ?  Tim B.
Tim is three steps away from John

John B.  Axel P.  Dan B.  Tim B.
“Three steps away” 😞 ?

Why Galaxy decided that this three steps away connection is a strong connection?
“Three steps away” 😞?

Why Galaxy decided that this three steps away connection is a strong connection?
Galaxy computes that this is a strong connection because of multiple ways of connections.
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

Dirk

Report

?
Galaxy provides a solution

Dirk

Project

Documentation

Ansgar

Report

Alexander Troussov “Mining Multidimensional Socio-Semantic Networks”

Guest lecture for students of Network and Distributed System MSc class in TCD 03.03.2008

© NEPOMUK Consortium 2006 – 2008
Dirk wonders “Why me?”
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 that which I don’t know that I want?
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 rankes these findings in a sensible way, depending on the topology of the cognitive map

• Whith Whom is Claudia connected?
  • With Dirk, Martin, Elaine, John, Hanna, etc?
  • With “some researchers”, like Dirk, Martin, Elaine, John, Hanna, etc?
  • With “many researchers”?
UI of Galaxy library
Generalisation
With Whom is Claudia connected?

All of these people

Claudia

Dirk

Martin

Elaine

John

Hanna

Researcher
Ranking

1 2 3
Ranking
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

Tag Assignments

Users

Resources

Tags

collaboration
connections
gos
innovation
koble
lotus portal
secondlife
and
soa
social-computing
socialnetworking
socialsoftware
software
swg
web20
forbiz
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

Instances

Timeline

Resources

Users

Instances

Java
JSP
Semantic Web
Industry

Tags

Reports to

Knows

Works with

Programming

C++
Java
Applets
JSP

Links to

Links to

Links to

Timeline

Users

Instances

Java
JSP
Semantic Web
Industry

Tags

Reports to

Knows

Works with

Programming

C++
Java
Applets
JSP

Links to

Links to

Links to
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 work on using 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
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, mezo, 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).
How to Find Out More

• 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
Thank you!

• Acknowledgements:
  • my IBM collaborators: Mikhail Sogrin and John Judge
  • Nepomuk Project Partners:
    • Digital Enterprise Research Institute - Brian Davis, Siegfried Handschuh
    • Cognium Systems - Alexander Polonsky and Mikhail Kotelnikov
    • SAP Research – Ying Du, Ernie Ong, Uwe Riss
    • DFKI - Leo Sauerman and Gunnar Grimnes
    • FZI - Heiko Haller and Max Voelkel