Social Semantic Desktop going Enterprise 2.0

Industrial Needs and Requirements to Use SSD on Enterprise 2.0 Intranet
Agenda

- IBM contribution to the Nepomuk project in a nutshell
- Enterprise 2.0
  - and IBM social software at a glance
- The use of SSD in a corporate environment
  - Specifics of corporate environment
  - and how this affects parts-and-parcels of SSD
- Research challenges (to make SSD useful on Intranet 2.0)
  - Automatic metadata generation / scalable backend for social software
- … and more details about IBM/Nepomuk approach to these challenges
  - IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks
- Road maps for exploration path within IBM and with IBM Enterprise 2.0
IBM/Nepomuk team results

- IBM/Nepomuk team efforts to improve consumability of social semantic desktop (SSD) in a corporate environment
- We focus on
  - Automatic metadata generation
  - Integration of SSD with corporate systems and data
- Narrow foci
  - Ontology based text processing to detect relevance of ontological concepts to a text (even if they are not mentioned)
    - to fully benefit from semantic web technologies you need to annotate your documents and emails with information as to what they are about (will you?)
  - Scalable and fast algorithms of graphmining
    - any SSD social applications (like community detection) needs to benefit from HUGE corporate knowledge (who reports to whom, geographical location of employees, employee profiles, …)
IBM/Nepomuk team results (cont)

- All IBM/Nepomuk software is available for download as one Java library codenamed Galaxy (http://www.alphaworks.ibm.com/tech/galaxy) (Galaxy library might be soon opensourced)
- Galaxy library is suitable for building micro-applications (components) for mining networked data
  - Soft clustering, fuzzy interencing, conceptual summary for a set of nodes on the network, Google Sets analogues, ... DEMO
- which can be used to build end-to-end solutions for various tasks of mining socio-semantic networks:
  - community detection, metadata recommendation, expertise location, ...
  - "food you might like", "you may know these people", ...
- What we can prove is that this “Lego car” (built from such micro applications) is easy to assemble, fast, scalable, and is a Mercedes for ontology-based text processing.
  - However, for many industrial applications easy to assemble and maintain “cars” are good enough
    - http://www.okino.com/tutorials/lego_import_and_rendering
IBM/Nepomuk team results (cont)

- Development of Galaxy is based on a new “mathematics”
  - Galaxy is based on Spreading Activation methods. Wikipedia reads: "Spreading activation is a method for searching associative networks, neural networks or semantic networks" and describes a particular algorithm of this method.
  - We are establishing the method(s) of spread of activation in their generic form, as efficient method for local search, relationship/association search, computing of dynamic local ranking, etc.
  - Developed the concept of fuzzy polycentric queries
  - We traced the methods of Spreading Activation back to methods of numerical simulation in physics
IBM/Nepomuk team results (cont)

- Galaxy is already IBM product for text processing as part of IBM LanguageWare
- We made insight into how to use Galaxy in several domains:
  - Including Ranking of Search Results, Related Item Recommendation, Modelling Temporal Aspects of Massive Dynamic Networks
  - Spreading activation is a search. Search for what? And how can I ask for what?
    - In mining socio-semantic networks “Its not what you know, its who you know”, “Its not who you know, its who they think you know”, “Its not who you know, its what they think you know” and so on. Therefore it is frequently not the search of nodes with particular properties, but the search of nodes with strong cumulative direct and indirect connections to the initial set of nodes. And activation spread is a very good method for this type of search.
  - Advantages of using spreading activation methods
    - No SPARQL or other explicit queries / No additional cognitive load / No browsing / Fast / Shows something of cognitive interest to perceive, contextualize, simplify, and make sense of otherwise complex interlinked data
IBM/Nepomuk team results (cont)

- Galaxy opens new opportunities for browsing and exploring data provided by networks
  - Working with KTH Human-Computer Interaction Group we developed new paradigm of many-to-many navigation similar to how one browses in a library or shop
  - This type of navigation is not simply browsing from a single object to another, but by dealing with several objects at the same time in a process similar to how one browses in a library or shop. Such an “ambient navigation” tool might be used for exploring various massive multidimensional networks, which occur in the socio-semantic information space which we encounter in the modern information age. This usage extends beyond the scope of scenarios explored in the EU 6th framework project Nepomuk where these ideas have been developed. For instance, Nepomuk Simple powered by Galaxy might be used to navigate not Nepomuk PIMO, but social networks.
Enterprise 2.0
Enterprise 2.0

- Enterprise 2.0
  - “Enterprise social software, also known as Enterprise 2.0, is a term describing social software used in "enterprise" (business) contexts. It includes social and networked modifications to company intranets and other classic software platforms used by large companies to organize their communication. In contrast to traditional enterprise software, which imposes structure prior to use, this generation of software tends to encourage use prior to providing structure” - Wikipedia

- Big, fast growing market
  - April 2008 report by Forrester Research is predicting that enterprise spending on Web 2.0 technologies is going to increase dramatically over the next five years, resulting in a global enterprise market of $4.6 billion by the year 2013

- Solutions from
  - Connectbeam, Facebook, Google (Blogger), MediaWiki, …
  - Two big solutions: Share Point (Microsoft) and Lotus Connections / Quickr (IBM):
    - “In short, Lotus is a better enterprise solution; Share Point is a better project or team solution”
What is Enterprise Web 2.0 (cont.)

http://www.readwriteweb.com/archives/enterprise_20_to_become_a_46_billion_industry.php
Multidimensional networks

- Ontologies
- Communication networks
- Web 2.0: Facebook, Delicious, LinkedIn, etc.
- Social networks
  - “… these networks become more and more multidimensional …” - Noshir Contractor
Delicious is a multidimensional Network

Tag Assignments

Users  Resources  Tags

- collaboration
- connections
- gos
- innovation
- kblue
- lotus portal
- secondlife
- social-computing
- socialnetwork
- socialsoftware
- software
- swg
- web20forbiz
Enterprise 2.0 (adds more dimensions)

This massive dynamic socio-semantic network is only a tiny part of Intranet 2.0 as it is shown on the next slide.

(Forget about mining methods which do not have (near) linear performance)
So, IBM Intranet is a multidimensional Network

IBM embraced Web 2.0 hits internally (Intranet 2.0) and made it into software for its customers.

Intranet Content

- **Fringe** ~ LinkedIn with people-tagging tool
- **Dogear** ~ Del.icio.us behind firewalls
- **Wikis, Blogs**
- **Beehive** ~ Facebook with new viral features

**PEOPLE**
aka
Employee
Multidimensional Networks in CI (Cyberinfrastructure) Multiple Types of Nodes and Multiple Types of Relationships

Prof. Noshir Contractor
Enterprise 2.0 – Why?

- Why Employer creates the hive and
- Why Employees create the honey, the buzz and provide pollination?
  (IBM’s Facebook-like is called Beehive, Beehive users are appropriately called bees)
  - Social tools provide a substitute for personal connections that flew away with globalization
    - In a global company (like IBM with nearly 400,000 employees), most people are too far away. “You cannot create a culture of innovation without creating a culture of collaboration – and at its core is creating a culture of trust with people you may never have met, and Web 2.0 tools help create trust.” Liam Cleaver, IBM
  - Facebook and the likes crush business productivity and hijack employees. Having Web 2.0 hits on Intranet is beneficial for attracting and retaining employees.
    - “Hotshots coming out of universities are accustomed to working across networks like Facebook and are likely to look at a company that still relies on the standard ’90s fare of e-mail and the phone as slow and backward”
    - “39% of 18 to 24 year-olds would consider leaving if they were not allowed to access sites like Facebook and YouTube”
  - Employees use Intranet 2.0 tools for collaboration, knowledge sharing and self-branding:
    - For instance, I created a “buzz” about the Nepomuk Summer School and provided useful “cross-pollination” by sharing this presentation with various IBM communities.
The use of SSD in Corporate Environments
Nepomuk SSD on Enterprise 2.0?

- If there is still a place to use the Nepomuk SSD on Enterprise 2.0?
- Yes – because the problem of knowledge representation in a formal, machine understandable manner problems is not yet addressed by Enterprise 2.0
- Semantic knowledge workbench, using semantic knowledge representation
  - Representing knowledge in a formal, machine understandable manner
  - Semantic resource annotation
  - Semantic inference
  - Semantic search and discovery: using ontological terms to describe a search
- where corporate knowledge is integrated with personal information models might help to transform Intranet 2.0 into a more personalized environment
Specifics of a corporate environment

- **High requirements to security**
  - Might prohibit the use of some Nepomuk SSD social features
    - Security has been a major barrier to peer-to-peer adoption.
      - Some IT professionals call peer-to-peer insecure, unscalable and unmanageable.
    - But peer-to-peer technologies are now supported by Intel and Sun peer-to-peer initiatives
      - so it might be the time to reevaluate peer-to-peer
  
- **“We know who you are”**
  - Simplifies identity management
Usefulness of SSD hinges on:

- Automatic metadata generation
- Software integration
  - For instance, Aperture plug-ins for Lotus Notes are needed
- Data integration
  - Corporate and personal knowledge integration
Corporate and personal knowledge integration

Social bookmarks are an example of corporate knowledge. Users of social bookmarking tool need to benefit from multiple connections.
Research Challenges of deploying SSD with Enterprise Web 2.0 Technologies

Two challenges IBM/Nepomuk team address
IBM Intranet 2.0 is a massive network

IBM embraced Web 2.0 hits internally (Intranet 2.0) and made it into software for its customers.

How can I navigate on this massive network? How can I find co-minded co-workers among 400+K employees?

Intranet Content

- Wikis, Blogs
- Beehive ~ Facebook with new viral features
- Fringe ~ LinkedIn with people-tagging tool
- Dogear ~ behind firewalls

PEOPLE
aka
Employee
Two Research Challenges

- **Automatic metadata generation** – important to improve consumability of SSD
  - EU 6th framework integrated project Nepomuk aims to build social semantic desktop based on Semantic Web standards. And to benefit from semantic web technologies you need to annotate your documents and emails with information as to what they are about. Automatic metadata generation must facilitate this tedious and error-prone manual process

- **Scalable, fast, small footprint backend for Social Software** is needed for usage of SSD in a corporate environment
  - Scalable
    - IBM community for WEB2 technologies has 1000+ members
  - Tolerant to errors, inconsistencies and incompleteness of data
    - IBMers update their profiles with average delay of 9 month
  - Social and Semantic processing needed to converge
    - the difference between social and “semantic” links in a corporate environment is blurred
IBM contribution to the Nepomuk project

Address these challenges:

- We created scalable graph mining technique
  - A blend of soft clustering and fuzzy inferencing
- We applied this to natural language understanding
  - Ontology-based text annotation
- We tested this technique for various scenarios of socio-semantic processing
  - Including related item recommendation
- are exploring the use of this technique for enabling SSD with “ambient” navigation
- We wrapped all this into one java library codenamed “Galaxy”
  - Available for download from IBM alphaWorks emerging technologies site
IBM Contribution: Text processing

- Galaxy as a product for semantic text analysis:
  - Reads the text
  - Memorises all the concepts
  - Uses networks of words to analyse which concepts sit well together

- Text which mentions *Mulhuddart, Lansdowne, Clontarf* is probably about *Dublin / Ireland / Europe*
IBM Contribution: Text processing (cont.)

- Galaxy builds semantic models of text documents as a function on nodes of a semantic network which shows the relevance of corresponding ontological concepts to the text
  - Vector Space Model (VSM):
    - Traditional Vector Space Model of Information Retrieval
  - Semantic Function Space Model:
    - Model we introduced which covers Vector Space Models and is somewhat similar to it,
    - however, VSM is an algebraic model, while Function Space Model can be studied by the methods of function analysis: find local extremes, make function “more smooth”, etc. involving graphmining
- Galaxy modifies this model to make it more similar to models of coherent and cohesive text as compared to models of random list of words
- The results show conceptual foci and disambiguate entities
Ontologies and their relatives (IBM/Nepomuk)

We pioneered new methods how to work with ontologies and semantic networks:

Our data are not “the truth, the full truth, nothing but the truth”, and some level of ambiguity is innate in social interaction. Fuzzy mathematics and soft computing can model very complex phenomena taking into account the inherently imprecise dimensions of nuanced empirical reality.
IBM Contribution: Socio-Semantic processing

- We tested this technique on multidimensional networks
  - which bring together people and all kinds of digital artefacts: documents, concepts, vocabularies, tasks, activities, and more.
- For different applications of socio-semantic processing
  - Collaborative filtering systems, where Galaxy can take into account not only relations entailed by instances of tagging, but also multiple connections between user, resources, tags (e.g. taxonomy driven folksonomy)
  - Large scale social network analysis
  - Model Temporal Aspects of Massive Dynamic Networks (synchronic and diachronic)
    - To provide structured view of massive networks on different levels of granularity: mega, mezzo, micro
    - To analyse trends
    - To anticipate recurring patterns of events

The picture – numerical simulation for the network of dozen of thousandths pixels in a grid delimited by a drawing of Henri Matisse
IBM Contribution: Nepomuk Integration

- Semantic text processing and navigation in semantic wikis
  - To automatically generate metadata for free texts
  - Integrated with iPad
- PIMO browser
  - Discover new relations in networks based on the strength of multiple connectivity between nodes
- Nepomuk Simple
  - IBM/Nepomuk component recommends related items for desktop “piles”
    - Like piles of things “Claudia needs to organise her trip to Belfast”
    - Or “piles” of shared objects in IBM Activity Explorer
  - Moving from isolated functionalities like “recommendations” towards comprehensive “ambient navigation” in socio-semantic networks
IBM component as a generic recommender

- Claudia is browsing her resources related to the CID projet meeting in Belfast
  - While Semantic Web technologies easily provide her with all the links to the resources explicitly connected to her task
  - Galaxy (IBM/Nepomuk component) reveals “hidden links”:
    - Galaxy locates and ranks all potentially useful resources (including people, flights, bus timetables) based on the analysis of networks related to Claudia’s Personal Management Model and corporate knowledge
- The screenshot shows related item recommendation in the Nepomuk PSEW
Road maps for exploration path within IBM and with IBM’s software
Road maps for exploration path

- We will use the Nepomuk SSD in IBM Ireland
- will investigate integration on software level:
  - E.g. writing Aperture plug-ins for Lotus Notes email client
- will investigate “semantic” integration
  - Integration of PIMO ontology and IBM corporate knowledge encoded in social networks, collaborative bookmarking system, blue pages etc
Read more:

- “IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks”
  - (http://www.alphaworks.ibm.com/tech/galaxy)

- “Navigation through the Set of Interlinked Resources via Polycentric Fuzzy Queries”
  - International SoNet Workshop 2008 (Camera ready copy will be available soon)

- "A Linguistic Light Approach to Multilingualism in Lexical Layers for Ontologies"
  - Computational Linguistics – Applications 2008 (Camera ready copy will be available soon)

- "Linguistically Light Lexical Extensions for Ontologies"

- “Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects”
  - Applications of Social Network Analysis Conference 2007

- "Ontology-based Text Processing for Personal Information Management"
  - Guest lectures at Universitat Autònoma de Barcelona in 2008

- "Mining Multidimensional Socio-Semantic Networks"
  - Guest lectures at Universitat Autònoma de Barcelona in 2008
    (http://atroussov.com/uploads/Mining_Socio-Semantic_Networks_Troussov.pdf)

- "Mining Socio-Semantic Networks Using Spreading Activation Technique“, 
  - Knowledge Acquisition from the Social Web, 2008
Research collaboration

- If the data in your task naturally lend themselves to be represented as a network, you might think about using Galaxy
  - Because Galaxy makes sense out of data provided by networks.
- To test Galaxy
  - Convert your network into the format which can be used by Galaxy (it is simple XML format)
  - ... and you’ll get first results in a matter of days
- To get good results
  - You might consider to work with us to leverage our experience with applying Galaxy to solve various problems
Thank you!