IBM Dublin Center for Advanced Studies & IBM LanguageWare

Mining Socio-Semantic Networks

Alexander Troussov,
IBM Dublin Center for Advanced Studies Chief Scientist
atrousso@ie.ibm.com

ESWC 2008, Tenerife, June 2008

Linguistically Light Lexical Extensions for Ontologies

- Lexical entries for many concepts are not that simple … often multi-word units. It seems that Linguistics is needed
  - Which linguistic theory do we use?
  - Can it be implemented for automation?
  - Who is going to do it?
    - Knowledge engineers who work with ontology data are not linguists
- IBM/DERI LEON (Lexical Extensions for Ontologies) and LLS (Linguistic Light Scanner):
  - Quick to implement
  - Flexible
  - Handling of multi-word and inflecting forms
  - Ease of use by non linguists

Linguistically Light Lexical Extensions for Ontologies (cont.)

- Lexical extensions drive OBIE, semantic annotation etc
- How it works:
  - **"… a nation united by its struggle … "**
    - SIGNATURE of **"United Nations"** is detected
      Variable word order, different capitalisation, and inflections are disallowed by constraints
      REJECTED
  - **"… consolidation of credit card debts …"**
    - SIGNATURE of **"debt consolidation"** is detected
      Variable word order is allowed, intervening tokens are allowed,
      inflected forms are allowed
      ACCEPTED
- Evaluation (our paper at Linguistic Resource Evaluation Conference 2008):
  - LEON+LLS gives a 4.9% increase in detecting multiword lexical items
    (before KE provides accurate descriptions)

Applications to semantic text processing

- Mapping from text to concepts in ontology (LEON-LLS)
- We build a semantic model of the text as function on nodes of a graph which represents a user's world view (their PIMO), or other ontology,
  - This shows how text is relevant to each of the concepts discovered.
  - We process this function taking into account empirics
  - This allows us to differentiate between semantic models of coherent and cohesive texts as compared to models of random lists of words
- Our Semantic Function Space Model of text
  - Somewhat similar to traditional Vector Space Model of Information Retrieval
  - However, VSM is an algebraic model, while Function Space Model can be studied by the methods of function analysis: find local maximums, make function more smooth, etc. involving graphmining

Applications to semantic text processing (cont.)

- Apply Galaxy reasoner to refine this modes using all those empirics which tell us the difference between random list of words and coherent cohesive texts
- Results (see Demo):
  - Keyword extraction (even if they are not mentioned in text)
  - Term disambiguation

Mining Multidimensional Networks

- Such as networks of concepts in an ontology, of social networks, or networks of people/organisations and the things they create and do
- Galaxy Library (http://www.alphaworks.ibm.com/tech/galaxy) reasoner used in semantic text processing:
  - Scalable
  - High Performance (200msc on huge networks)
  - Provides elements of soft clustering and fuzzy inferencing
  - Superficially, Galaxy UI performs like Google Sets: User focuses on one or more concepts, and Galaxy tries to predict other concepts of interest.

Mining Multidimensional Networks (cont.)

- Advantages of using Galaxy Library
  - 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?

Applications to egocentric queries in socio-semantic networks

- For instance, as a people finder, Galaxy library works like a radar:
  - Locates people of potential interest to you
    - co-workers, friends-of-friends, ...
    - co-bloggers, co-taggers, ...
    - co-co-workers, co-co-bloggers, co-co-taggers, ...
    - taggers-of-co-bloggers, bloggers-of-co-taggers, ...
    - ...
  - Ranks them according cumulative strength of direct and indirect socio-semantic relations with you
- Learn more:
  - “Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects”, Applications of Social Network Analysis 2007 (ASNA 2007, September 2007, Zurich)

Applications to egocentric queries (cont.)

- Galaxy 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*”?

Polycentric queries

- In applications to social tagging systems (like Del.icio.us or IBM's dogear)
- All the data can be viewed as a network, where users, resources and tags are related by instances of tagging
- And, by mining such networks, Galaxy library can provide all "traditional" functionalities like
  - Community detection
  - Community-based tag recommendation
  - Expertise location
  - etc
- Taking into account additional multiple relations:
  - Like relations between people,
  - Relation that the tag JSP might be related to the tag Java
  - Hyperlinks between resources

Ambient Navigation

- Ability to perform efficient polycentric queries open the way for “Ambient Navigation”:
  - 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
- For instance, in applications to social tagging:
  - User can get metadata recommendations for a particular resource or a group of resources, get explanations, locate experts, etc

Learn more:

- From my recent guest lectures at Universitat Autònoma de Barcelona:
  http://atroussov.com/activities
- IBM LanguageWare Miner for Multidimensional Socio-Semantic Networks manual
  http://www.alphaworks.ibm.com/tech/galaxy
- “Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects”, Applications of Social Network Analysis 2007 workshop, September 2007, Zurich
- “Linguistically Light Lexical Extensions for Ontologies”, International Conference on Language Resources and Evaluation (LREC), May 2008 , Marrakesh