Discovering Tag Semantics for Ontology Building, Enrichment & Population in Multilingual Environments

Andrés García-Silva[†]

† {hgarcia}@fi.upm.es
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Boadilla del Monte, Madrid, Spain

- Introduction
- Problem and Objectives
- Discovering tag semantics
  - State of the Art
  - Proposed approach
    - Multilingual Sense Repository
    - Sem4Tags
- PhD progress and future work.

- Social Tagging Systems
  - Web 2.0 applications
  - Applications for storing, sharing, and discovering information resources.
  - Users assign *tags* in multiple languages to identify information resources
  - *Tags* are used to search/discover resources
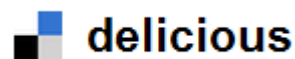
- **Folksonomy**
  - Emerging classification scheme from social tagging systems
  - Folk: People, Taxonomy: Classification
  - Represented by: **Users, Tags, Resources**
  - Two **Tags** are related if:
    - They co-occur when annotating a **resource**
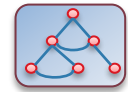   - They co-occur when a **user** use them
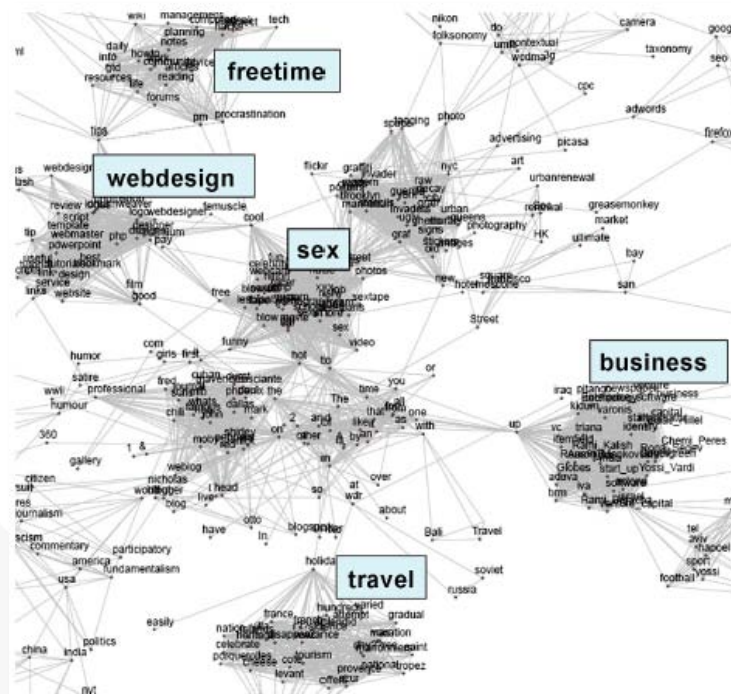
- ## Folksonomies & Ontologies

  - We can use folksonomy information to create ontologies or to add knowledge to existing ontologies.

    - New classes and relations

    - New instances

Folksonomy → Ontology

Advantages

- Knowledge generated by the masses
- Are low cost:
  - » Users are keen to assign tags
  - » Low cognitive burdens
- Evolve fast:
  - » new vocabulary is continuously added by users
- Vocabulary tends to converge

Drawbacks

- No fixed vocabulary
- Lack of semantics associated with tags
  - Ambiguous tags
  - Relations among tags are not defined (sameAs, broader-narrower, subclassOf, typeOf)
- Multilingual Tags

- **Problem**: How to take advantage of multilingual folksonomy information to build ontologies, and to enrich, and populate existing domain ontologies.
  - **Discover tag semantics in multilingual folksonomies**
  - How to use tag semantics to build ontologies, and to add knowledge to existing folksonomies.

- **Objectives**:
  - Methods and techniques to:
    - Extract knowledge from multilingual folksonomies
    - Use that knowledge to build new ontologies, and to add new classes, relations and instances to existing domain ontologies.
  - Software to implement the proposed methods and techniques.

Approach	Type	Auto	Data Source	Data Selection & Cleaning	Context Identi-fication	Disambi-guation	Semantic Identi-fication
Mika, 2007	Clu	Yes	Delicious, other	Yes	Yes	No	Yes
Hamasaki et al., 2007	Clu	Yes	Polyphonet	No	Yes	Yes	No$^{\mp}$
Giannakidou et al., 2008	Hyb	Yes	Flickr	Yes	Yes	Yes	No$^{\mp}$
Jäschke et al., 2008	Clu	Yes	Delicious, BibSonomy, Other,	Yes	Yes	No	No$^{\mp}$
Specia & Motta., 2007	Hyb	Semi	Delicious, Flickr	Yes	Yes	Yes	Yes
Angeletou et al., 2008	Ont	Yes	Flickr	Yes	Yes	Yes	Yes^{+}
Cantador et al., 2008	Ont	Yes	Union of Flickr and Delicious	Yes	Yes	No	Yes^{+}
Tesconi et al., 2008	Ont	Yes	Delicious	Yes	Yes	Yes	Yes^{+}
Passant, 2007	Ont	No	Enterprise Folksonomy	Yes	Yes	Yes	Yes^{+}
Maala. et al., 2007	Ont	Yes	Flickr	Yes	Yes	No	Yes

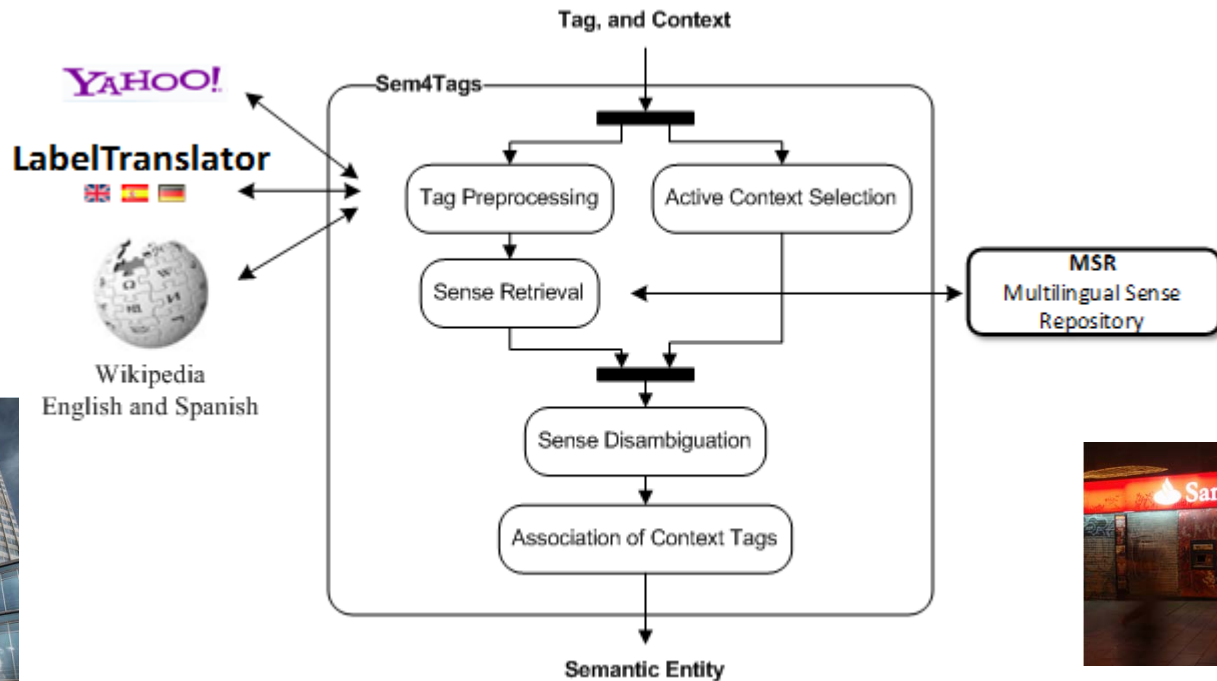**None of the analyzed approaches deals with multilingual tags**

- **MSR:** a Multilingual Sense Repository based on Wikipedia and enriched with semantic information taken from DBpedia.

- **Sem4Tags**: A process for Associating of Semantics to Tags.

YAHOO!

LabelTranslator

Wikipedia
English and Spanish

Tag, and Context

Sem4Tags

Tag Preprocessing | Active Context Selection

Sense Retrieval

MSR Multilingual Sense Repository

Sense Disambiguation

Association of Context Tags

Semantic Entity

Europe
Euro
Finance
Central bank
awesomePic
Nikon ..

Bank → http://dbpedia.org/resource/Bank ← Banco

Dinero,
Calle,
Santander,
Money,
Madrid,
Atm,
cajero

- First Year
  - Paper accepted in the Knowledge Engineering Review.
    - State of the Art: Discovering and Associating Semantics to Tags in Folksonomies
  - Research visit to the IAM group of Southampton University
  - Paper accepted in the Workshop CKCaR'09 (Collective Knowledge Capturing and Representation)
    - Preliminary Results in tag Disambiguation using DBpedia.

- Second Year
  - Paper sent to the *EKAW-2010*
    - Associating Semantics to Multilingual Tags in Folksonomies.
  - Send a paper to the *ISWC-2010* (end of Jun)
  - Research visit (*Jul-Dec*)
  - Send a paper to the *Artificial Intelligence Journal, Special Issue on "Artificial Intelligence, Wikipedia and Semi-Structured Resources" (end of Oct)*
  - Send a paper to a workshop describing the proposed process for ontology building, enrichment & population from folksonomy information

- Third Year
  - Send a paper to a conference describing the evaluation of the proposed approach for ontology building, enrichment and population from folksonomy information.
  - Write PhD Thesis
  - Research Visit