Relating Folksonomies with Dublin Core

Maria Elisabete Catarino
Ana Alice Baptista
Universidade do Minho
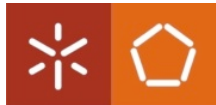
Universidade do Minho

CAPES

BRASIL
UM PAÍS DE TODOS

**What
this presentation is
about**

- Contextualization - KoT – how the project began and the first indications

- The research project - what we are willing to know

- Tag analysis - rules and procedures

- Results: of the pilot study and some preliminary final results
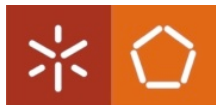
- What we'll do next

Universidade do Minho

How this project began

- Liddy's post on DC-Social Tagging mailing list;

- Preparation of a proposal and posting it to the mailing list;

- Receiving expressions of interest from people from the UK, Spain, France, Belgium, Italy and USA;

- A communication on the DC Social Tagging workshop at DC2007 with contributions from: Ana Alice Baptista, Emma L. Tonkin, Andrea Resmini, Seth Van Hooland, Susana Pinheiro, Eva Mendéz and Liddy Nevile.

Universidade do Minho

Some Indications from the first phase

- Users apply tags not only to **describe the resource**, but also to describe their **relationship with the resource** (e.g. to read, to print,...)

- **Tags don't always correspond to atomic values** - Many of the tags have more than one value, which potentially results in more than one metadata element assigned.

- **The majority of tags could be mapped to DC properties, but many were left unmaped** - 14 out of the 16 DC elements, including Audience, have been allocated.

- **Four new properties were broadly identified** - "Action Towards Resource" (e.g., to read, to print...), "To Be Used In" (e.g. work, class), "Rate" (e.g., very good, great idea) and "Depth" (e.g. overview).
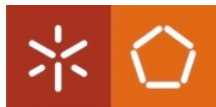
Universidade do Minho

However....

- the first phase consisted in very preliminary work just to give us an idea of what kind of results we could expect;

- a more detailed work would be needed in order to have more accurate results

- challenge to a PhD student: Elisabete.

Universidade do Minho

The questions

- Do DC properties have the necessary semantics to clarify and express how given tags relate to the resource to which they were applied?

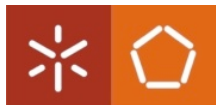If not, which other properties that hold this semantics can be identified to complement DC and to be used in social tagging applications?

Universidade do Minho
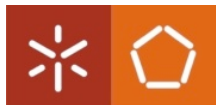
The Data Set

- The Data Set
  - 5098 tags
  - 15381 users
  - 75429 Tag occurrences

- The whole study was made **manually** in order to be as precise as possible regarding the meaning of the tags

The organization of the study

- Pilot study - 5 resources; 355 users; 311 tags; 1141 tag occurrences.

- Full study – 50 resources; 15381 users; 5098 tags, 75429 tag occurrences

- Stages:

    1. Analysis of Tags;

    2. Identification of complementary properties;

    3. Formalization of the new properties in a ontology;

    4. Validation by the community

Universidade do Minho
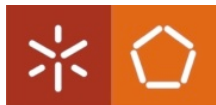
Tag Analysis - procedures

- Analysis of all tags in the data set;

- If the meaning of tags is not clear:

  - to use lexical resources (dictionaries, encyclopedias, WordNet, etc)

  - analyse other tags of the same users

  - Contact the user

- Grouping of tags into key-tags

  - Key-Tag is a normalised tag that represents a group of similar tags

Universidade do Minho
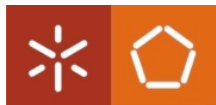
Analysis of Tags: rules

- Only tags written in Latin alphabet were considered;

- English was the chosen language to represent Key -Tags.

  - Tags written in different languages, but, most of the tags were written in english;

Universidade do Minho

Analysis of Tags: rules

- Compound tags related to only one concept = only one key-tag

  – Example:

    – **Tag**: `Institutional Repositories`
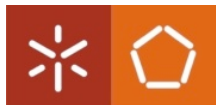
    Modifier   Focus

    – **1 Key-Tag**: `Institutional Repositories`

Analysis of Tags: rules

- Compound tags related to two or more concepts = two or more key-tags.

    - Example :

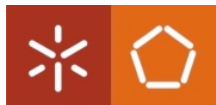   – **Tag**: `Classification-Cataloguing`

        Focus   Focus

        – **2 Key-Tags**:   `Classification`
                             `Cataloguing`
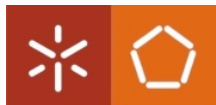
Analysis of Tags: rules

- When Simple Tags could clearly be post-coordinated they were analysed as a compound term.
  - For instance a user who, when assigning tags to the resource "The Semantic Web" written by Tim Berners-lee, insert the following tags:
    - `the, semantic, web, article, by, tim, berners-lee`

The system of Social Bookmarking generated seven tags. However, it is clear that these tags can be post-coordinated to have a meaning such as:

> *Title* – The Semantic Web
>
> *Creator* – Tim Berners-Lee
>
> *Subject* – semantic web

Universidade do Minho
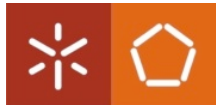
Analysis of Tags: rules

- Tags may correspond to more than one properties
  - Example - Simple Tag:
    - `Architecture`
      - Properties: **Title** and **Subject** (resource title: "An Architecture for Information")
  - Example - Compound Tag:
    - `Doi:10.1045/april2002-weibel`
      - Properties: **Identifier** `doi:10.1045`; **Date** `april 2002`; **Creator** `Stuart L. Weibel`

Universidade do Minho

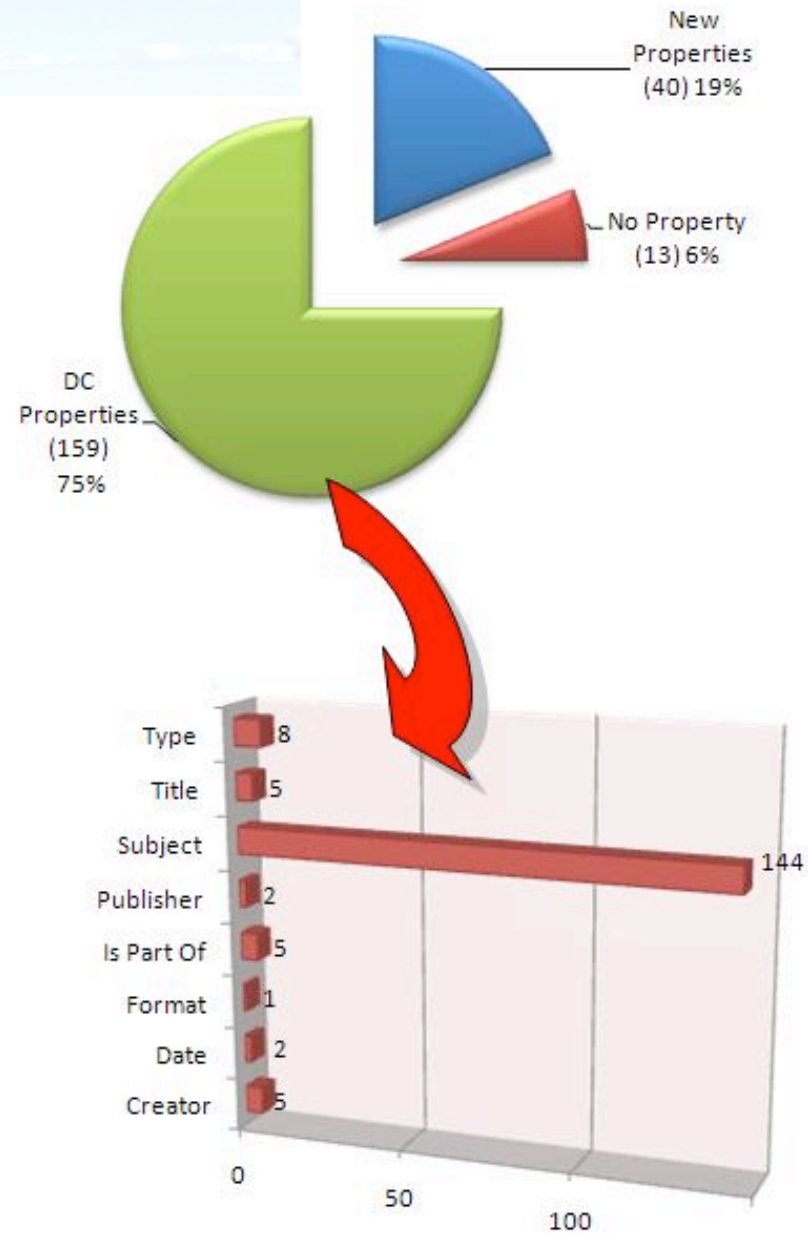# *Complementary properties - procedures*

- Identify potential new properties related with the tags that couldn't be assigned do DC properties

- These are specific for Social Tagging applications.

Universidade do Minho

PILOT STUDY Results

New Properties (40) 19%

No Property (13) 6%

DC Properties (159) 75%

- 5 resources
- 212 Key-tags
- **159 (75%) – DC Properties**
- 40 (19%) – New Properties
- 13 (6%) – No property

Type 8
Title 5
Subject 144
Publisher 2
Is Part Of 5
Format 1
Date 2
Creator 5

0   50   100

Universidade do Minho

PILOT STUDY
Results

- 5 resources
- 212 Key-tags
- 159 (75%) – DC Properties
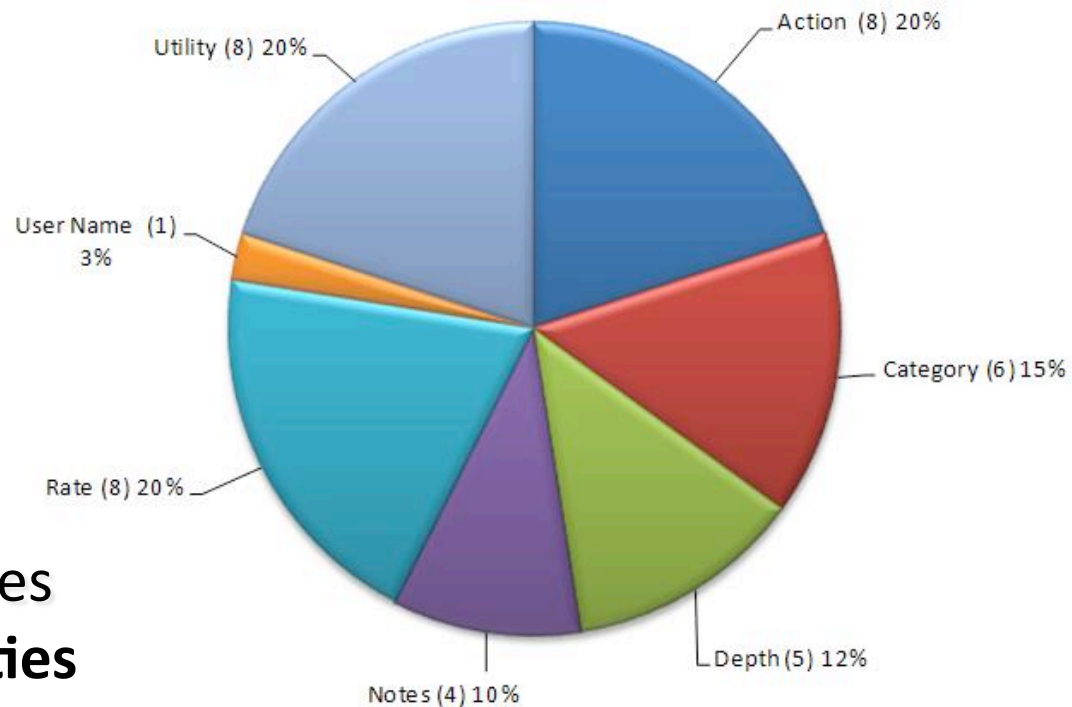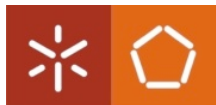- **40 (19%) – New Properties**
- 13 (6%) – No property

Utility (8) 20%
Action (8) 20%
User Name (1) 3%
Category (6) 15%
Rate (8) 20%
Depth (5) 12%
Notes (4) 10%

Universidade do Minho

Potential New Property
Action

- Action of the user in relation to the tagged resource.

- Example:

```
toread
_toread
a_lire
```

Universidade do Minho

Potential New Property
Category

- This property includes tags whose function is to group the resources into categories.

- Example:
  - Key-Tag `DC Tagged` it was noticed that the corresponding resources had also other tags woth the prefix **dc:** (e.g.: *dc:contributor, dc:creator, dc:publisher, dc:language*, among others)

Universidade do Minho

Potential New Property
Depth

- This type of tag confers the degree of intellectual depth to the tagged resource.

- Example:

  `Diagram, doc/intro, overview, semanticweb.overview, semwebintro.`

  - These tags was applied for only one resource, means that users are describing a resource which content is thought as a schematic or a summarized explanation, introductory and general.

Universidade do Minho

Potential New Property
Notes

- This property may be proposed to represent the tags that are used as a note or reminder.

- Example:
  - Tag `Hey`, refers to Tony Hey*, a well-known researcher who made a debate on important issues that were related to the tagged resource.

  this information was given by the user who assigned the tags

Universidade do Minho

Potential New Property
Rate

- Rate, meaning pattern, category, class or quality is important to include tags that are evaluating the tagged resource.
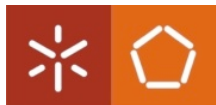
- Example:

```
great
good
old
```

Potential New Property
User Name

- The property "User Name" labels the resource with the name of tagger himself.

- Example:

  `Alttablib`

Universidade do Minho
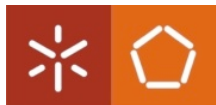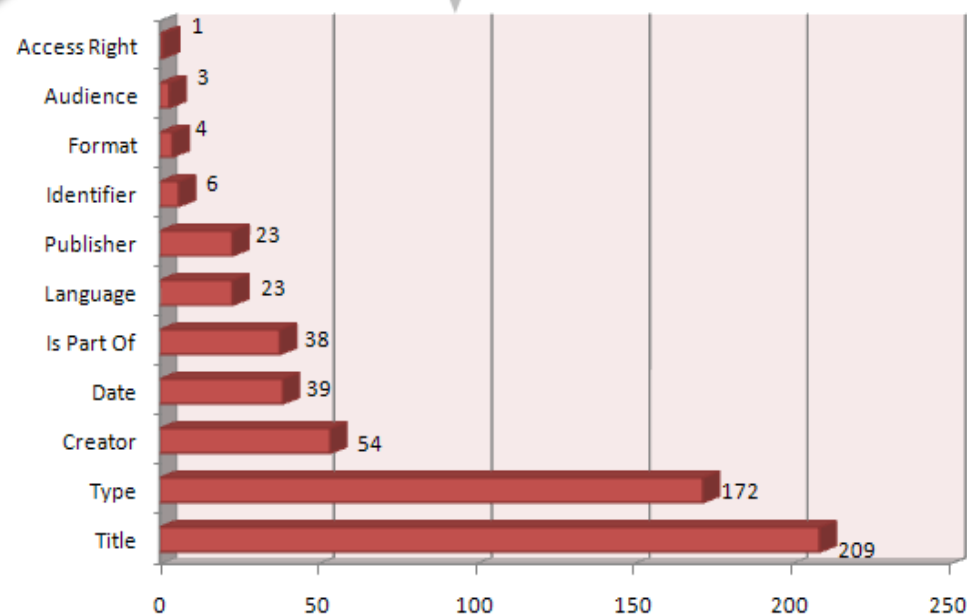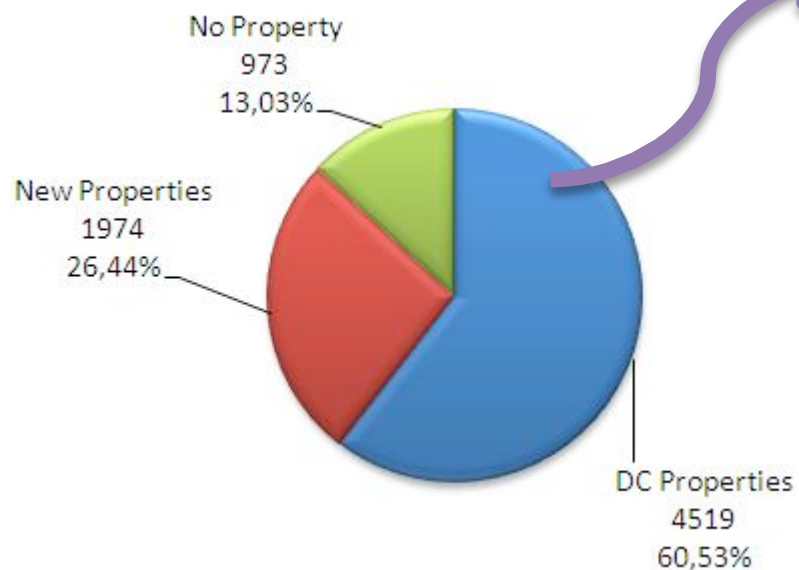
Potential New Property
Utility

- This property refers to the utility of the resource for the user.

- Example:
  - `Maass`, is a tag was bundled in "study". This term represent the name of a teacher, information found in the user's notes: *Forschung von Prof. Maass na der Fakultat Digitale Medien na der HFU*
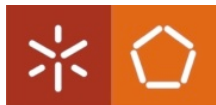
Universidade do Minho

Preliminary results from the final study: DC Properties

Subject; 3947; 87,34%

Other DC Properties; 572; 12,66%

No Property 973 13,03%

New Properties 1974 26,44%

DC Properties 4519 60,53%

Access Right	1
Audience	3
Format	4
Identifier	6
Publisher	23
Language	23
Is Part Of	38
Date	39
Creator	54
Type	172
Title	209

0   50   100   150   200   250
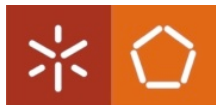
Universidade do Minho

Conclusions

- DC properties can be assigned to a great part of the tags analyzed in the pilot study and in the final study (75% and 60% respectively)

- However, still, a significant number of tags them are left out (40% in the final study)

- 26% of the tags in the final study can be assigned to potential new properties specific for Social Tagging Applications

- The pilot study indicates that there 7 potential new properties and the final study points out to 10.

What we'll do next and how you can help

- Validation of the study and feedback: questionnaires will be handed out at the social tagging workshop and will also be available in survey monkey (more info will be sent to you via email);

- Proposal of an application profile for social tagging applications.

- Develop partnerships for an international project in the semantic Web context that applies this application profile (contact us if you are interested injoinning)
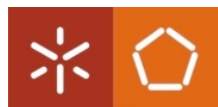
Universidade do Minho

Thanks!!!

Elisabete Catarino - ecatarino@dsi.uminho.pt

Ana Alice Baptista - analice@dsi.uminho.pt