Towards a Knowledge Diversity Model

Rakebul Hasan, Fabian Flöck, Katharina Siorpaes, and Reto Krummenacher

Slides: Denny Vrandečić and Elena Simperl

DiversiWeb @ WWW2011

March 28, 2011, Hyderabad
Collaboration on the Web is broken

Dieses Wiki hat ein Problem

Wikipedia and diversity

- Lemma selection level
  - Do you want an article for each and every „The Simpsons“ episode?
    - Depending on your decision
      - your audience will change
      - your author pool will change
      - the perception of your project will change

- Is the number of „members of Muslim faith“ identical in
  - The Arab language Wikipedia
  - The Hebrew language Wikipedia
  - The English language Wikipedia

- Will borders be described differently in various Wikipedia language editions?
Diversity support for wikis

Chocolate

Table of contents showToC.toggle("show","hide")

1 Different kinds of chocolate
2 The history of chocolate
3 Chocolate as a stimulant
4 Why chocolate tastes so good
5 How chocolate is made
6 Chocolate in the media
7 External links

Different kinds of chocolate

Chocolate is a common ingredient in many kinds of sweets -- one of the most popular in the world -- made from the fermented, roasted, and ground seeds of the tropical cacao tree Theobroma cacao. Dictionaries refer to this cacao substance as "chocolate," which is an intensely flavored bitter (not sweet) food, although this is legally defined as cocoa in many countries. This is usually sweetened with sugar and other ingredients and made into chocolate bars (the substance of which is also and commonly referred to as chocolate), or beverages (called cocoa or hot chocolate).

Chocolate is often produced in the form of little sculptures, for example as rabbit- or egg-shaped chocolates, near a holiday in many countries called Easter.

Dark chocolate
Milk chocolate
Semisweet chocolate (used for cooking purposes)

The history of chocolate

The Aztecs associated chocolate with xochiualcatl, the goddess of fertility. In the New World, chocolate was consumed in a drink called xocoatl, often seasoned with vanilla, chili pepper, and nopalte. Xocoatl was believed to fight fatigue, a belief that is probably attributable to the caffeine content. The drink was said to be an acquired taste. José de Acosta, a Spanish Jesuit missionary who lived in Peru and then Mexico in the later 16th century, wrote:

Loathsome to such as are not acquainted with it, having a scum or froth that is very unpleasant to taste. Yet it is a drink very much esteemed among the Indians, where with the feast noble men who...
Diversity-enhanced Web

Diversity-enabled Web applications

Diversity-enabled collaboration technology
- Wiki platform
- Blogging platform
- Micro-blogging platform

Formal model for diversity information and information retrieval algorithms
- Models, lightweight reasoning
- Search, selection and ranking
- Summarization
- Presentation and interfaces

Algorithms for mining diversity from text
- Opinion detection
- Multilinguality
- Bias in media
- Fact coverage
- Story links

Data management infrastructure
- Blogs
- News
- Twitter
- Wikipedia
- LOD
- Annotation, integration, linking
Diversity model

- Captures all notions related to diversity-enabled information management in a machine-understandable ontology
- Based on established upper-level ontologies and models in information science
- Core concepts
Knowledge diversity glossary (paper)

- Agent
- Belief
- Bias
- Data
- Diversity
- Emotion
- Entity
- Event
- Fact
- Information
- Information object
- Knowledge
- Metadata
- Object
- Objectivity
- Object feature
- Opinion
- Opinion expression
- Opinion holder
- Polarity of opinion
- Sentiment
- Subjectivity
- Text
- Topic
Core concepts

- **Agents** holds **opinions** on a topic
- **Topics** can be everything
- A **Document** is an information object containing **opinion expressions**
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of **opinions expressed** by the **agent** or in the **document**
- **Diversity** is the co-existence of **biases** for a **topic**
Core concepts

- **Agents** hold opinions on a topic
- **Topics** can be everything
- A **Document** is an information object containing opinion expressions
- **Opinion** and opinion expression
- **Bias** of an agent or document is the set of opinions expressed by the agent or in the document
- **Diversity** is the co-existence of biases for a topic

- **Examples**
  - editor of wiki article
  - journalist
  - publisher
  - blogger...

- Agents are associated with additional metadata
  - basic metadata
  - social network
  - history of publication
Core concepts

- **Agents** holds opinions on a topic
- **Topics** can be everything
- A **Document** is an information object containing opinion expressions
- Opinion and opinion expression
- Bias of an agent or document is the set of opinions expressed by the agent or in the document
- Diversity is the co-existence of biases for a topic

Examples
- of a wiki article
- of a news story

Topics can be related to each other and aggregated

Choice of a reference topic ontology would lead to biases
- No commitment to one topic ontology
Core concepts

- **Agents** holds **opinions** on a topic
- **Topics** can be everything
- A **Document** is an information object containing **opinion expressions**
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of **opinions** **expressed** by the **agent** or in the **document**
- **Diversity** is the co-existence of **biases** for a **topic**

- **Example**
  - article
  - news story
  - tweet
  - blog post…
- Documents are associated with basic metadata
- Diversity mining algorithms operate on documents to identify opinions
Core concepts

- **Agents** holds **opinions** on a topic
- **Topics** can be everything
- A **Document** is an information object containing opinion expressions
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of opinions expressed by the agent or in the **document**
- **Diversity** is the co-existence of biases for a topic

- **Opinion expression** is the actual representation of an opinion in a document
- **Beyond** simple sentiment analysis
  - Not just positive, negative, neutral
  - e.g. “Palestine is a country”
Core concepts

- **Agents** holds opinions on a topic
- **Topics** can be everything
- A **Document** is an information object containing opinion expressions
- **Opinion and opinion expression**
- **Bias** of an agent or document is the set of opinions expressed by the agent or in the document
- **Diversity** is the co-existence of biases for a topic

- Bias lies in the selection of opinion expressions
- Algorithms can be devised that predict a bias from previous history of opinion expressions and to calculate relationships between biases
Core concepts

- **Agents** holds **opinions** on a topic
- **Topics** can be everything
- A **Document** is an information object containing **opinion expressions**
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of opinions expressed by the agent or in the **document**
- **Diversity** is the co-existence of **biases** for a **topic**
Example: Chocolate

Bias
Chocolate is healthy

Opinion
Chocolate lowers cholesterol

Agent
User: Equinox

Topic
Chocolate

Opinion expression
“Chocolate lowers cholesterol”

Document
Wikipedia: Chocolate

Diversity
Opinions on healthiness of chocolate

author
contains
expresses
holds
requires
defines
about
about
Next steps

- Representation in OWL
- Grounding in existing ontologies
  - DOLCE, SKOS, Dublin Core, SPAR
- Fill the concepts with life / data
- Definition of APIs to access diversity-enriched data
vielfältigkeit
éagsúlacht
разнообразие
diversidad
Agents hold opinions on a topic

Topics can be everything

A Document is an information object containing opinion expressions

Opinion and opinion expression

Bias of an agent or document is the set of opinions expressed by the agent or in the document

Diversity is the co-existence of biases for a topic