



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



WIKIPEDIA
Die freie Enzyklopädie

Dieses Wiki hat ein Problem

Entschuldigung. Diese Seite hat momentan Schwierigkeiten.
Warte einige Jahre und versuche dann neu zu laden.

Du könntest in der Zwischenzeit

Conservapedia
The Trustworthy Encyclopedia

search

Go Search

popular links

- Main Page
- Educational Index
- Debate Topics
- Recent changes
- All pages
- All images
- All logs
- Random page

help

- Help Index
- Statistics
- Conservapedia Commandments

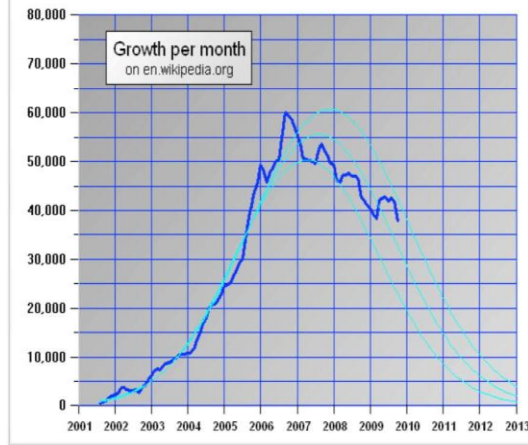
world history

- World History Lectures
- World History Homework

edit console

What links here

Date						
2002-01-01						
2003-01-01						
2004-01-01	440 000	415 000	188 000	176 000	47 000	61 000
2005-01-01	1 400 000	1 600 000	447 000	620 000	189 000	315 000
2006-01-01	3 300 000	6 000 000	901 000	2 800 000	350 000	719 000
2007-01-01	6 200 000	10 600 000	1 500 000	4 600 000	541 000	949 000
2008-01-01	9 500 000	10 900 000	2 200 000	4 300 000	710 000	817 000
2009-01-01	12 300 000	11 700 000	2 700 000	4 400 000	873 000	868 000
2009-07-01	13 700 000	11 400 000	3 000 000	3 800 000	928 000	681 000



Declining workforce participation is another factor affecting the Output Gap since the



In the News

Liberal Hypocrisy: NPR's Long War against Fox News, Conservatives.

NPR and parent Public Broadcasting System may pick and choose which of its employees' opinions are considered firing offenses, but there is one thing that remains consistent about the so-called "nonprofit news organization" — its bias against conservatism and the Fox News Channel. [4]

"Liberal Logic" Watch: Bad Driver? In Debt? NYC Says Give Up Your Guns!

If you live in New York City and want to own a gun, a new law would make it more difficult — if you are a litterbug, have a poor driving record or are behind on bills. Under proposed revisions to the police department's handgun, rifle and shotgun permit procedures, the NYPD can reject gun license applicants for a number of silly reasons including:

"If they have been arrested or convicted of almost any "violation," in any state; having a "poor driving history"; having been fired for "circumstances that demonstrate lack of good judgment"; having "failed to pay legally required debts"; being deemed to lack "good moral character," or if any other information demonstrates "other good cause for the denial of the permit."

What is it about liberals that render them incapable of reading the Constitution clearly? [5]

Very Unfavorable Ratings for Harry Reid, Nancy Pelosi Hit New High. With midterm congressional elections just a week away, the number of voters who view Speaker of the House Nancy Pelosi and Senate Majority Leader Harry Reid Very Unfavorably have reached their highest levels yet, especially among the critical Independent Voters. [6]

Attention Silly Liberals: Only 14% Prefer Government-Regulated Economy Over Free Market. Likely voters

No Weapons Allowed

ATTENTION CRIMINALS
This is a Defense Free Crime Zone
All Law Abiding Patrons Of This Establishment
Have Been Disarmed For Your Convenience
Enjoy
The Management



- 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"
wikipedia

COLOR

group

individual

text changes

text age

SPACING

date

versions

Daniel Quinlan "removing obscure heraldry information, belongs on [[heraldry]] if anywhere"

authors posts

Dmerrill 1
Larry_Sanger 3
Lee Daniel Crocker 1
Conversion script 1
Oliverkroll 1
Brion VIBBER 1
Bth 1
Oytolan88 1
Dachshund 1
JakeVortex 1
Olivier 1
KF 1
Vera Cruz 1
Magnus Manske 1
Zanimum 1
Elywa 2
Youssef'san 1
Wik 1
Evercat 1
Cyp 1
Rmhermen 1
Daniel Quinlan 1

Chocolate

(Revision as of 18:32, 6 Aug 2003)

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**.

Table of contents showToggle("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 an extremely popular ingredient, available in many types, and great quantity. Different forms and flavors of chocolate are usually produced by varying the amount of the ingredients used to make the chocolate.

Derk chocolate

Milk chocolate

Semisweet chocolate (used for cooking purposes)

The history of chocolate

The *Aztec* associated chocolate with *Xochiquetzal*, the goddess of fertility. In the New World, chocolate was consumed in a drink called *xocolatl*, often seasoned with *vanilla*, *chili pepper*, and *pimento*. Xocolatl was believed to fight fatigue, a belief that is probably attributable to the caffeine content. The drink was said to be an acquired taste. *Jose 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 they feast noble men who

Diversity-enhanced Web



Google™



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

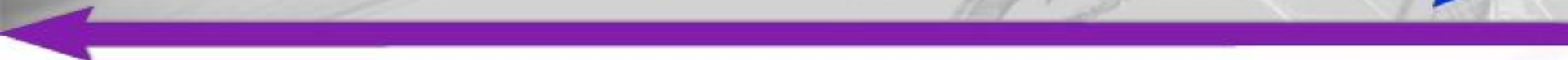
Annotation, integration, linking



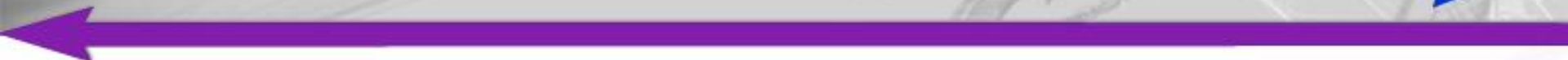


- Captures all notions related to diversity-enabled information management in a machine-understandable ontology
- Based on established upper-level ontologies and models in information science
- Core concepts



- 
- 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



- 
- **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**



- **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
 - editor of wiki article
 - journalist
 - publisher
 - blogger...
- Agents are associated with additional metadata
 - basic metadata
 - social network
 - history of publication



- **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



- **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



- **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”*



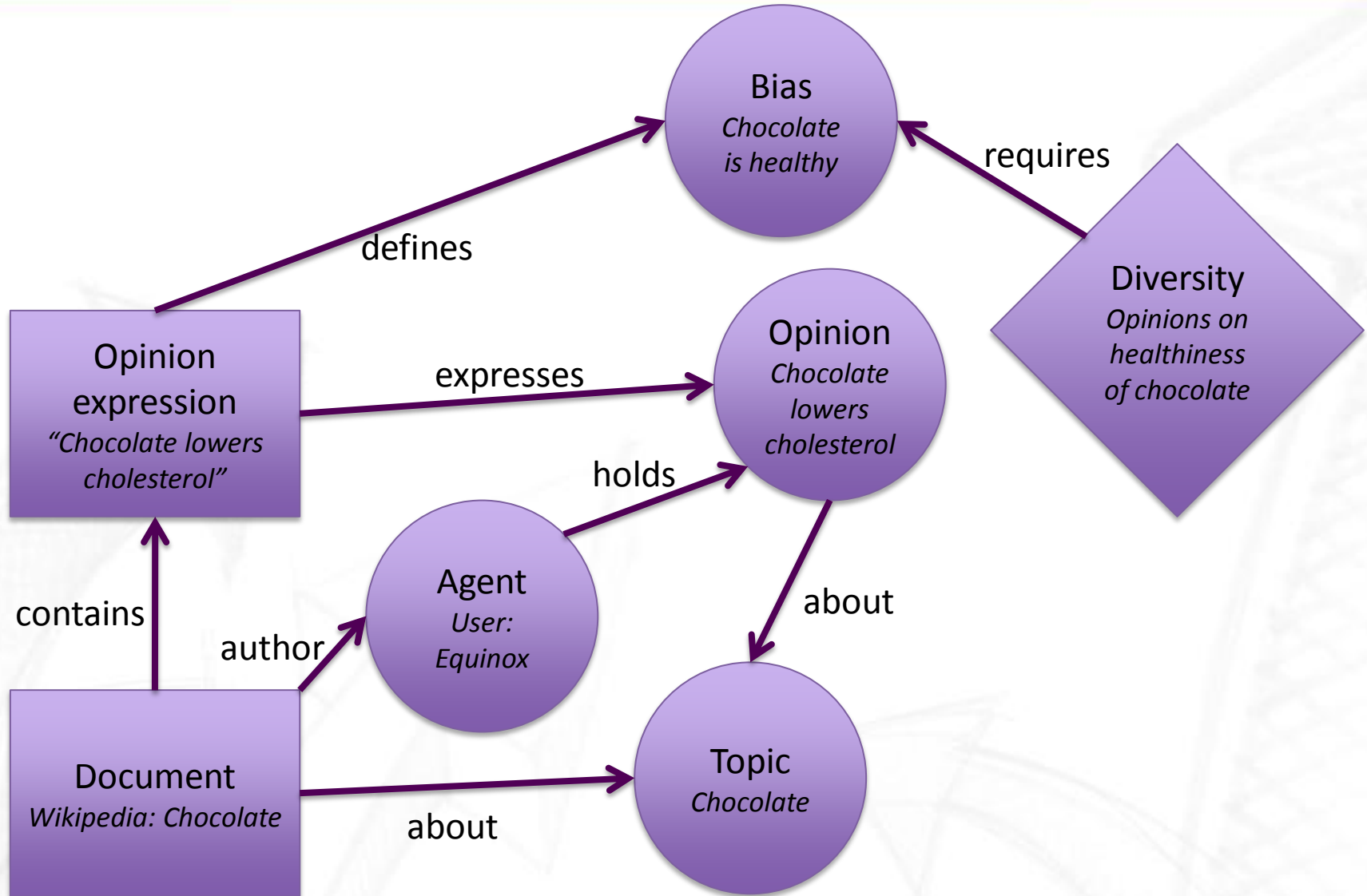
- **Agents** holds **opinions** on a **topic**
- **Topics** can be everything
- A **Document** is an information object containing **opinion expressions**
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of **opinions expressed** by the **agent** or in the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



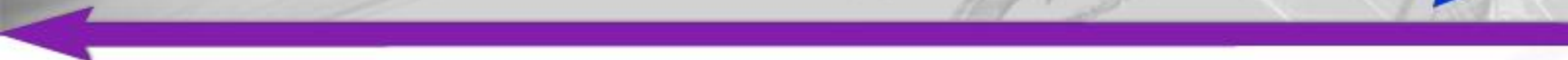
- **Agents** holds **opinions** on a **topic**
- **Topics** can be everything
- A **Document** is an information object containing **opinion expressions**
- **Opinion** and **opinion expression**
- **Bias** of an **agent** or **document** is the set of **opinions expressed** by the **agent** or in the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



Example: Chocolate





- 
- 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



mångfald diversidade
 variatie **vielfältigkeit** różnorodność
 rozmanitost
 éagsúlacht diversité
 monimuotoisuus manifoldighed
 πολυμορφία **разнообразие**
 įvairovė diversitată sokszínűség
diversidad diversitate
 mitmekesisus



- **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**

