Diversity (in Media)

prof.dr.ir. Arjen P. de Vries
arjen@acm.org

Living Web 2009 (Invited Talk)
The First International Workshop on Living Web: Making Web Diversity a true asset; collocated with the 8th International Semantic Web Conference ISWC-2009

Supported by the FP7 FET European Project: Living Knowledge - Facts, Opinion and Bias in Time
How it started...

World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists, Policy, November's W3 news, Frequently Asked Questions.

What’s out there?
- Pointers to the world's online information, subjects, W3 servers, etc.

Help
- on the browser you are using

Software Products
- A list of W3 project components and their current state. (e.g. Line Mode, X11 Viola, NeXTStep, Servers, Tools, Mail robot, Library)

Technical
- Details of protocols, formats, program internals etc

Bibliography
- Paper documentation on W3 and references.

People
- A list of some people involved in the project.

History
- A summary of the history of the project.

How can I help?
- If you would like to support the web.

Getting code
- Getting the code by anonymous FTP, etc.
How it looks today...

Peter Bailey  Bing ships Bing Twitter aka Social Search - fantastic to see us shaking things up, being agile and shipping product, not press releases. Delighted my team - especially Krishna - could help out.

Bing
Source: www.bing.com

Yesterday at 04:48  ·  Comment  ·  Unlike  ·  Share

You and 4 others like this.

Arjen P. De Vries  I like the document + evidence organisation of the interface! Almost like expert finding at TREC-ENT :-)  
Yesterday at 08:26  ·  Delete

Write a comment...
Diversity

- Diversity ← Data Creator
Jane Winans: Maybe I'll become a vegetarian.
Jane just found out her Death's Fortune in Mobsters 2.
Date: December 21, 2009 at 9:30 am
Caused: Impacted by a meat hook.
Yesterday at 15:40 via Mobsters 2: Vendetta Comment Like
[
Find out when you'll die]

Recent Activity
- Jane just completed the Peevish Muscle for a Lean Shark mission in Mobsters
  Comment Like
- Jane and Michele Purington are now friends.
  Comment Like
- Jane and Laurie Cummings are now friends.

Jane Winans: Like the looks of these... hope they help
Jane sent a set of Glock 17 pistols to Jane Smith and Justin Talker using Mobsters 2: Vendetta Gifts.
Yesterday at 05:17 via Mobsters 2: Vendetta Gifts Comment Like
Send Mobsters 2 Vendetta Gifts to me (Jane)

Jane Winans: And miles to go before I sleep...
Yesterday at 04:50

View all 4 comments

Ian Cameron: Dec 21
Yesterday at 20:46

Jane Winans: I see. That explains a lot right there, lol.
Yesterday at 21:17

Write a comment...

Jane Winans: HA... that is cool! I love Calvin!

Jane completed the quiz "What Calvin alter ego are you?" with the result Tracer Bullet.
The detective created by Calvin's fantasy. With no friend but your gun, you're just way too cool.
Yesterday at 04:25 via Quiz Monster Comment Like
Take this quiz

Green Turquoise likes this.
Write a comment...
Arjen P. De Vries: Nice SIGIR paper from Jun, looking beyond the Probabilistic Ranking Principle.

Portfolio Theory of Information Retrieval
Source: webis.de/acta...
Portfolio effect of Document Ranking. As a general rule of thumb, when we present search results for a query, documents should...

Jun Wang: I am very pleased you guys like the Ideal I intend to increase the awareness of using variance and higher moments for IR modeling. Hopefully more comes later.

11 hours ago  Delete


11 hours ago  Delete

Jun Wang: The “Suboptimal” paper has a fair argument on PRP and has been the starting point of my paper.

11 hours ago  Delete

Mark Sanderson: There was also a keynote talk at SIGIR 1999 by an economist. Have you seen that? Looking back, one can see that talk was a call for diversity in ranking (most of us ignored that talk).

11 hours ago  Delete

Jun Wang: A very good point. In fact, the early attack comes from Cooper himself, although it never explicitly mentioned. 


http://www.spadec.com/EDOC/.../read text.../recordDetail.../read more...

11 hours ago  Delete

Mark Sanderson: Also, didn’t the Robertson PRP paper explore reasons why it might not work? I think I heard Steen say that he discussed such possibilities in the original paper.

11 hours ago  Delete

Jun Wang: That, although the PRP was first mentioned and examined critically by Cooper in his unpublished technical report - Robertson did mention this in his PRP paper. Unfortunately, the technical report is not online - I must get a hard copy from Cooper recently. It is always interesting to look at the history of IR.

11 hours ago  Delete

Mark Sanderson: Just to digress, it sounds to me as if the Cooper tech report should be added to the SIGIR digital library that Donna Humes is organizing.

11 hours ago  Delete

Jun Wang: Let me ask Cooper first to get his comment...

11 hours ago  Delete

Fernando Díaz: Jun, you may want to cite this Leks paper in particular (you don’t). Since large sections of your paper seem to be founded on it.

7 hours ago  Delete

Jun Wang: Check the first sentence of related work section 5. It may not be the same paper but same idea from the same author. Sheddie anyway.

6 hours ago  Delete
Re: Living Web – CFP

- Diversity underlying content creation process

+ 

- Success of Social Web applications

= 

- Difficult to get a balanced view on a given subject using current search technology
User-defined relevance criteria

- Many studies (from 1966 on) suggest relevance criteria beyond topicality
- Evaluation of relevance is closely tied to the requester's situation and goals
  - experience, cognitive state and perceptions
- Empirical study: Schamber 1991
  - How do users judge weather information services?
  - 811 mentions of different criteria, clustered into 10 criteria
Schamber 1991

- Accuracy
- Currency
- Specificity (detail, concreteness)
- Geographic proximity
- Reliability or reputation
- Accessibility
- Verifiability
- Clarity of presentation
- Dynamism (ability to interact)
- Affective response
Barry 1994

- Finds similar criteria “mentioned by users evaluating the information within documents as it related to the users' information need situations”
- User factors like experience, background, knowledge level, beliefs, personal preferences
- Evaluations of individual documents take place within the larger context of the information environment
(continued: Barry 1994)

- The user's ability to judge the reputation of the author depends upon being presented with information about the author

- Explore possible connections among relevance criteria and specific portions of document representations
User-defined relevance criteria in web search

- Savolainen and Kari, 2006
  - Predictive judgement (on link)
    - Accepting: specificity, topicality and familiarity
    - Rejecting: insufficient accessibility and lack of specificity
  - Evaluative judgement (on page)
    - Accepting: specificity and topicality
    - Rejecting: lack of specificity, insufficient accessibility
User-defined relevance criteria in image/video search

- Choi and Rasmussen, 2001
  - Topicality
  - Time frame
  - Accuracy

- Yang and Marchionini, 2004
  - Topicality
  - Genre
  - Duration
  - Cinematography
  - Motion
~ user-defined relevance criteria

However... taking success into account:
  – Technology that more systematically looks into diversity, helps identifying bias and different viewpoints

Naïve users could accept only results from their own biased view – may make them happy, but not us!
  – A paternalistic retrieval system?!
  – Is Living Web technology usable?!
Diversity

- Diversity ⟷ Data Creator
- Diversity ⟷ Data Consumer

Social Media:
- Consumer becomes a co-creator
- Data 'consumption' traces
Blog Rockin' Beats:
The 21-year-old Kyteman (trumpet player, composer and Producer Colin Benders), has worked for 3 years on his debute: the Hermit sessions.

Kluun.nl:
De site van Kyteman

Jazzenzo:
... a performance by the popular Kyteman's Hiphop Orkest

Ticket Sales Luxor theatre:
May 22nd - Kyteman's hiphop Orchestra - www.kyteman.com
Diversity

- Social Media:
  - Consumer becomes a co-creator
  - Data 'consumption' traces

- In essence: many new sources to play the role of anchor text!
Marc Davis  Tim Berners-Lee basically said AOL is to WWW as Facebook is to interoperable social graphs #w2s.
8 hours ago  Comment  Like

Marc Davis  Tim Berners-Lee talking about how the WWW was originally focused on enabling collaboration and shared linking and annotation #w2s.
8 hours ago  Comment  Like
TWITTER extra.

**Most recent Tweets about #w2s**

- **yager:** "Only 20-25% of humanity actually uses the Web at all" Tim Bemers-Lee #w2s (via @graw) - 17 minutes ago

- **jord:** RT @graw: RT @benghoster Tim Bemers-Lee HTML 5 is something to be excited for, will cause web pages to turn into widgets #w2s - 17 minutes ago

- **Ischalke:** RT @austinhill: "Collecting data is less important than connecting people" - Sean Parker #w2s (via @austinhill) - 37 minutes ago

- **Christopher:** RT @austinhill: Loved Grammy's breakdown of the mantra panel, which was a total highlight of #w2s http://bit.ly/2zV3yER (austinhill) - 38 minutes ago

See more Tweets about #w2s

**Top links shared in Tweets about #w2s**

- **ReadWriteWeb Interview With Tim Bemers-Lee, Part 1: Linked Data**
  - readwrite.com/archives/interview_with_tim_bemers_lee_part_1.php

- **Paul and Partners:** #timbl "I don't see why SPARQL servers shouldn't provide CSV files" http://is.gd/4xZTX (readwrite.com) vi

- **RNC:** In case you missed it in July, ReadWriteWeb Interview With Tim Bemers-Lee http://bit.ly/QuLwU (readwrite.com) (lots of great content in that) #w2s - 8 hours ago

See more Tweets about readwriteweb.com/archives/interview_with_tim_bemers_lee_par...

- **Brian Ahler - Health IT & Healthcare Reform: Web 2.0 Summit - Marissa Ma...**
  - ahler.blogspot.com/2009/10/web-20-summit-marissa-mayer.html

- **ehealthpr:** RT @bogus: Watch Marissa Mayer at the Web 2.0 Summit announce the Google/Twitter deal & Google Social Search http://bit.ly/UG4kC (blogplickr.com)

- **Andrewspong:** This creeps me out, for some reason RT @ahier: M. Mayer at the Web 2.0 announcing Google/Twitter deal http://bit.ly/UG4kC (blogplickr.com) (via @crappily) - 1 hour ago

See more Tweets about ahier.blogspot.com/2009/10/web-20-summit-marissa-mayer.html

- **WikiRecent Questions including "Who is Tim O'Reilly?"**
  - wikipedia.com/Q/FAQ/1038

- **Hagelotest:** Which questions on @Answersdotcom can you answer about @morelly http://omly/QOY9 (answers.com) & @johnbullock http://omly/QO0y #w2s - 1 hour ago

See more Tweets about wikipedia.com/Q/FAQ/1038

See more Tweets about @Answersdotcom
Live TV from Web 2.0 Summit here with select sessions starting tunes, 10/20 @ approx 4:30 pm Pacific

Watch today's @Web2Summit closing keynote with Tim Berners-Lee & five* starting at about 5pm PDT http://bit.ly/1MGl1JJ #w2s jsa 16 hours ago by @crelymedia

ReadWriteWeb Interview With Tim Berners-Lee, Part 1: Linked Data

In case you missed it in July: ReadWriteWeb Interview With Tim Berners-Lee http://bit.ly/Ct7EU (lots of great context in that) #w2s 8 hours ago by @nwm

WikiAnswers - Tim O'Reilly Questions including "Who is Tim O'Reilly"

#w2s Over 100 questions on Tim O'Reilly http://is.gd/1HgU1 and John Battelle http://is.gd/4vug2 on Answers.com. See how many you can answer? 17 hours ago by @magicaula

Early Access Release Channels (Chromium Developer Documentation)

eXSite: Sergey Brin says we can use Chrome for Mac by going to http://bit.ly/DUjks (expand) pre-beta.

#Chrome #w2s 11 hours ago by web2summit

WikiAnswers - John Battelle Questions including "What is one thing John Battelle loves about his wife?"

#w2s Over 100 questions on Tim O'Reilly http://is.gd/1HgU1 and John Battelle http://is.gd/4vug2 on Answers.com. See how many you can answer? 17 hours ago by @magicaula

Web 2.0 Summit Blog » Blog Archive » Web 2.0 Summit Recap - What's Launched

Web 2.0 Summit Recap - What's Launched (and videos of sassists!) http://bit.ly/35v9tM From the Web 2.0 Summit blog #w2s 15 hours ago by @w2s

The very first web page was mroc01.cern.ch/hypertext/WWW/TheProject.html see it archived here

RT @LTimesNystrom: The very 1st website can be found here: http://j.mp/TpQ4e #w2s via @NataliaW 8 hours ago by @tchagner
#w2s!
PetaMedia NoE

The Community (user2user)

- Community-based incentive
- Shared infrastructure and costs

The P2P Net

- CrowdSourcing
- Shallow/incomplete semantic input

User network

Physical Network

Data Network

Based on
- Content
- Usage
- ...

PetaMedia’s main focus
Social Video Dataset

- Develop a **common data-set** that is large enough to allow meaningful research, and contains video content *as well as* explicit information generated by a **social network of sufficient density**
- [irp5petamedia.pbworks.com](http://irp5petamedia.pbworks.com)
Requirements

- **Availability of video data**
  Creative Commons (CC) licenced data only

- **Availability of social data**
  Users creating the data should be organized in a social network, and provide feedback about their preferences in relation to the data (comments, ratings, ...).
Blip.tv (25-60% CC)
Metadata/annotations (but... insufficient)

About this show:
O'Reilly Where 2.0 Conference
Since 2004, the Where 2.0 Conference has been the place where the grassroots and leading edge developers building location aware technology intersect with the businesses and entrepreneurs seeking out location apps, platforms, and hardware to gain a competitive edge. In the O'Reilly conference tradition, Where 2.0 presents leading trends rather than chasing them.

Details
Genre: Technology
Language: English
Content Rating: TV-UN

Contact
Website:
E-mail: Not publicly available.
AOL IM: Not provided.
Skype: Not provided.
Y! Messenger: Not provided.

Cast
O'Reilly Where 2.0 Conference hasn't shared any cast information with blip.tv yet.
Extend metadata/annotations

1) Join Blip.tv with Del.icio.us and Twitter
2) Crawl links to videos and social networks of users creating those links

Up to 4 levels of social network, only 2 levels of metadata (bookmarks/posts/profiles)
Social metadata/annotations

- 14K distinct links to blip.tv (from 22K bookmarks) in TU-Berlin DAI Labor del.icio.us collection (2003-2007)
- 27K links to blip.tv in Topsy
Lawrence Lessig, on Palin's inexperience - http://blip.tv/file/1287551 a twelve minute, v
Probably old to most, but it was new to me. Lessig on Palin: http://blip.tv/file/1287551
Lawrence Lessig on Palin's experience: http://blip.tv/file/1287551
retweeting @kplawver: Lessig's take on Palin's Experience: http://blip.tv/file/1287551
Wow. Lessig methodically tears it up: http://blip.tv/file/1287551
Retweet @Torrie: This is amazing. Everyone should watch it. http://blip.tv/file/1287551
This is amazing. Everyone should watch it. http://blip.tv/file/1287551
Food for thought: An interesting piece called "Sarah Palin's Experience in Just 12 Minute
Holy cow. Side by side comparison of Palin's experience vs. all VPs. Worth every second
palin's experience in just 12 minutes. http://blip.tv/file/1287551
Link: Palin's experience in just 12 minutes http://tinyurl.com/3j3e2w This presentation
12 min history lesson on Vice Presidential experiences: http://blip.tv/file/1287551 very
Is Sarah Palin experienced enough to be Vice President? splendid analysis here: http://
todo mundo fala sobre os keynotes do SJ. Mas pra mim o rei do keynote Ã© o larry less
Sarah Palin's experience in 12 min, incredible presentation by Lessig: http://blip.tv/file/
Excellent video by Lessig http://blip.tv/file/1287551 on the experience of Palin compare
Lessig compares Palin's experience to all other VPs: http://blip.tv/file/1287551 (via @x
Watching Lawrence Lessig's 'Palin's experience in just 12 minutes' http://tinyurl.com/3j
how did this video not make the fp of Digg or Reddit? http://blip.tv/file/1287551
Tweets about blip.tv

- E.g.: http://blip.tv/file/2168377
  - Amazing
  - Watching “World’s most realistic 3D city models?”
  - Google Earth/Maps killer
  - Ludvig Emsgard shows how maps/satellite pics on web is done (learn Google and MS!)
    - and ~120 more Tweets
Diversity surfaces

- Fermented garlic? Interesting! (via @ChefMark) Chef Mark's Quick Bites: Black Garlic  
  http://blip.tv/file/2313261
- @ChefMark Love learning new stuff from Culinary Media Network. Good stuff from recent show  
  http://blip.tv/file/2297478 yummmm
- why would anyone want to stop you from voting?  
  http://blip.tv/file/1401303
- great video about making your vote count (music by DJ Spooky):  
  http://blip.tv/file/1401303
- Want to see Brussels at Christmas watch this video  
  http://shadowandjames.blip.tv/file/1656052/
Every search request is metadata!

That same metadata proves useful as training data for multimedia content analysis!

‘Police Sting’
Sting performs with The Police

‘Elton Diana’
Sting attends Versace memorial service

‘Led Zeppelin’
Sting performs at Led Zeppelin concert
Diversity from User Logs

- Image Search Clickthrough Data
  - Many query variations
  - Reliable (multiple studies, including our own; even over 80% in some cases)

- Method:
  - Present different query variants' clicked images in clustered view
  - Merge different query variants' clicked images in a round robin fashion into one list (CLEF)
ImageCLEF 'Olympics'

Olympic games

Olympic torch

Olympic village

Olympic rings

Olympic flag

Olympic Belgium

Olympic stadium

Other
ImageCLEF 'Tennis'
ImageCLEF Findings

- Many queries (>20%) without clicked images
  - Corpus and available logs originated from different time frame
- Best results combine text search in metadata with image click data for topic title and each of the cluster titles
- Using query variants derived from the logs increases recall with 50-100%
  - However, also topic drift; reduced early precision
More diversity?

- Preliminary results:
  - Map queries into linked data cloud, find 'abstract' patterns, and re-use those for query suggestion, e.g.:
    - A and B play-soccer-in-team X
    - A is-spouse-of B
- Right mix of statistics and semantics?
Presents search results grouped by facet.

Presents answer types grouped by facet.

Additional information upon mouse-over on answer-type.
Case for Uniformity?

- **Vox Populi**
  - Stefano Bocconi, Frank Nack, and Lynda Hardman. Vox populi: a tool for automatically generating video documentaries, ACM Hypertext and Hypermedia, 2005

- **Vox Populi generates video documentaries based on interviews about controversial topics**

- **User selects topic and a point of view**
Views on the war in Afghanistan?

I am not a fan of military actions.

I cannot think of a more effective solution.

War has never solved anything.

Two billions dollar bombs on tents.
Two billions dollar bombs on tents

I cannot think of a more effective solution

I am not a fan of military actions

War has never solved anything
Children as Users

- Leave out extreme views?
  - *Especially* for adolescents?
Age vs. Reading Level
Summary

- Topicality is 'just' one of many relevance criteria
  - Others include specificity, currency, verification, affectiveness, ...

- However, as opposed to this workshop, studies into user-based relevance criteria usually have not considered search success as a factor
  - What about the user who is happy with biased results?
Summary

- Diversity surfaces in information use
  - PetaMedia corpus
  - User log-based image retrieval
- Show information sources together with their use context
  - Opinions will show; group as facets?
- Do not just rank complete objects
  - Media fragments
  - Entities
Vlaams Radio Orkest performing in the rotunda of the Ravenstein Gallery
Simon Blackley, flickr.com/people/sblackley/
Vlaams Radio Orkest performing in the rotunda of the Ravenstein Gallery
Simon Blackley, flickr.com/people/sblackley/
Vlaams Radio Orkest performing in the rotunda of the Ravenstein Gallery
Simon Blackley, flickr.com/people/sblackley/