



Opinion mining and sentiment analysis

W3C SEMINAR, JUNE 19TH 2012

David Osimo, Tech4i2.com

www.crossover-project.eu

[#pmod](#)



THE CHALLENGES

- Making sense of a thousand voices across different platforms
- Identifying “good ideas”
- Reduce uncertainty on possible impact of policies by getting early and real-time feedback

POSSIBLE SOLUTION

Argument mapping software helps organising in a logical way these policy statements, by explicitating the logical links between them..

Voting Advise Applications help voters understanding which political party (or other voters) have closer positions to theirs.

Automated content analysis helps processing large amount of qualitative data.

STATE OF THE ART



ShareThis



TRACKING OPINIONS ON TWITTER

twitrratr

SEARCH

SEARCHED TERM

digitalagenda

POSITIVE TWEETS

12

NEUTRAL TWEETS

64

NEGATIVE TWEETS

0

TOTAL TWEETS

76

15.79% POSITIVE



rt @digitalagendaeu: a great video explaining what's #opendata <http://t.co/u32kkf5t> - you too, share your video w/ the #digitalagenda family <http://t.co/da8k4twv> (view)



rt @digitalagendaeu: a great video explaining what's #opendata <http://t.co/u32kkf5t> - you too, share your video w/ the #digitalagenda family <http://t.co/da8k4twv> (view)

84.21% NEUTRAL



EU consults on Internet of Things (IoT) <http://t.co/ynmX4qS0>. GCHQ to identify all things on net in new Project IdIoT. #DigitalAgenda (view)

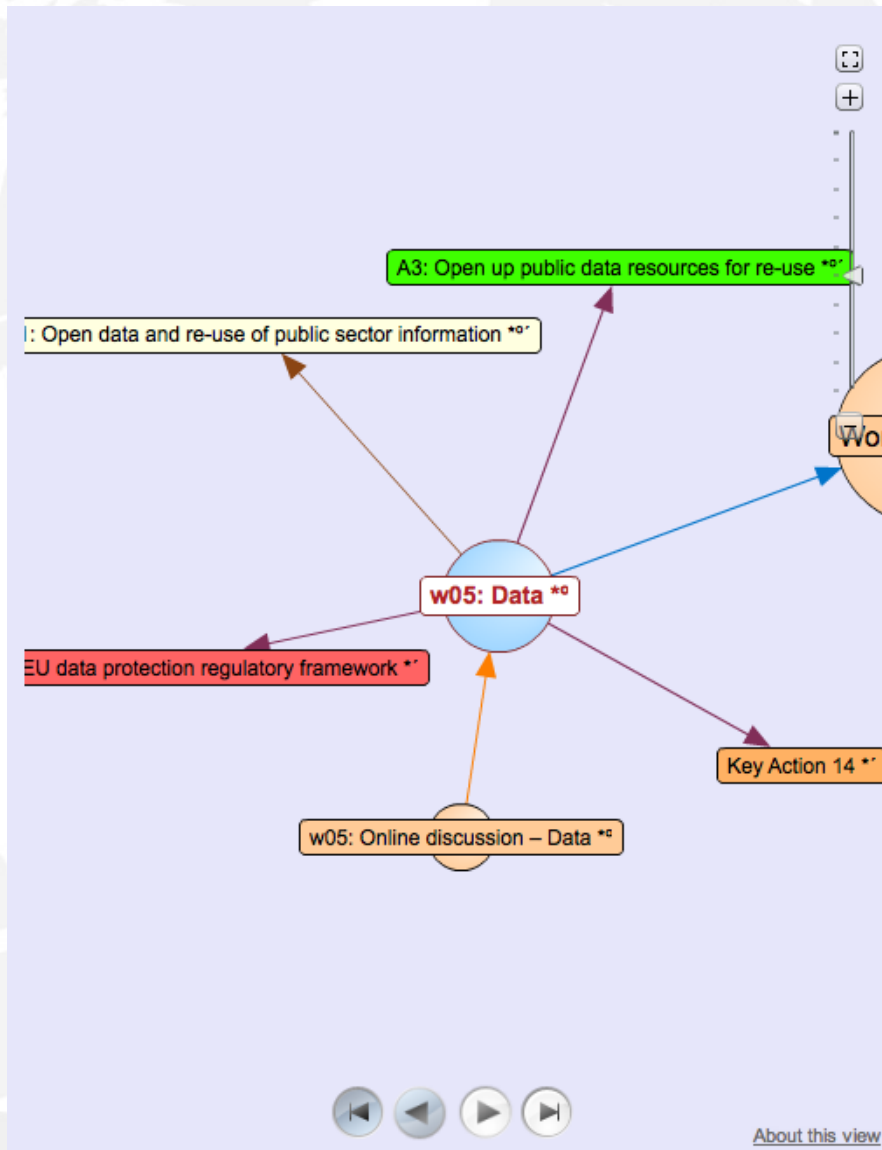


The #EU Commission starts consults on rules about connectivity of "Internet of Things" <http://t.co/Qca1vOUS> #DigitalAgenda #Europe2020 (view)

How can businesses be fully prepared and take advantage of

0.00% NEGATIVE

None Found



w05: Data Position #152995

The workshop will explore the potential of data, some of the most promising economic and business aspects involved, and discuss how policy for data and our investment in R&D can better address the challenges of businesses and the public sector and further support innovative business development.

Recent industry reports detail how data volumes are growing rapidly, with annual increases up to 40%. Data volumes of Facebook are by themselves estimated to be as big as the entire World Wide Web was in 2004.

Exploiting this data - extracting actionable meaning out of the mass of unorganised and often volatile material in many languages, analysing, using and re-using it- could transform the Web and the way we handle information once and for all. It could facilitate our lives through innovative ICT applications, lead to better support for decision-making and planning, give research in different domains a major boost, and improve interaction between online, social and enterprise usages. In this scenario, the ability to draw knowledge out of data would change the Internet in a way similar to the changes brought about by the 'search' and 'social media' revolutions in recent years.

Programme

1. Opening keynote speech focussing on the business opportunities of data.
2. Three consecutive panels covering specific application areas, each fire-starting with "testimonials" followed by immediate discussion on the opportunities, hurdles and issues. The selected areas constitute a large component of data in both volume and economic value:

I Public Sector Information (PSI) in different application areas such as geographic information (with possible cooperation with the JRC), business registry (with possible cooperation with DG MARKT), transport, etc.

Discussion should focus on issues such as charging, licensing, changing mindset of Public Sector Bodies (PSB) and competition for SMEs that comes from PSB that produce their own PSI-based products. It should further facilitate an exchange of views on the steps which could be taken to overcome road blocks and foster openness

Questionnaire > Customize your voting advice > Voting advice

Questionnaire

Support smartvote!

I value smartvote: Flattr 163

1. Welfare state & family (0/4)

Categorized 1-by-1

1. Do you support an increase in the retirement age for both women and men (e.g. to 67)?



yes	rather yes	rather no	no	No answer	Weight
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	=

2. Would you welcome the introduction of a "degressive" daily allowance for unemployment insurance (i.e. the daily allowance would decline as the period of unemployment grows)?



yes	rather yes	rather no	no	No answer	Weight
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	=

Information

- To receive a voting advice you must answer at least one question.
- The more questions you are answering, the more precise your voting advice will be.
- If you need additional information regarding a specific question, please click on the corresponding information button. You will receive additional information on the topic of the questions as well as pro and contra arguments.
- You can also weight your answers.

ideas for ...

2 people from UserVoice are here



I suggest you ...

- enter your idea (new feature, fix bug, etc) -

add

[top ideas](#) [new](#) [completed](#)

43 votes for

OpenID user authentication planned

... would allow those sites who use OpenID to directly correlate their users with their suggestions, and other neat stuff.

5 comments | by [brycoman](#) | is it? [spam](#) [duplicate](#)

28 votes for

make it possible to "star" suggestions planned

I would like to be able to "star" certain suggestions that I want to remember (if I have no more votes but want to keep tabs on interesting suggestions).

comment | by [torhengb](#) | is it? [spam](#) [duplicate](#)

18 votes for

Allow a members only site

I would like to have a site that only members that I approve can view the suggestions. The application has many competitors who would likely come to view the suggestions and steal them if they were available to everyone.

8 votes left!

How good ideas are born and raised:

I suggest you ...
add more95 votes
How many votes?
1 2 3
make other

submit

vote up

marcus
openid is a great so

discuss

official response
Not sure how we could d
open to ideas in the core
- [comment](#)

response

bole completed
that connection is tight our fo
demand and as the "Star" user - it
business and about it or more for
done!



UserVoice activity feed

Want your own UserVoice page?

ADVANCED TOOLS

Opinion Space: Silverman Research Help | Sign Out

[Show more suggestions](#) [Show Top Suggestions](#) [Show who I've rated](#) Suggestions Rated: 5

participant1855 ✕

[View Opinions](#)

How could Unilever improve its IA policies?

This is where participants write their response to the discussion question. We are looking for suggestions to improve the topic under investigation.

[Flag for review](#)

How relevant is this suggestion to you?

Not Relevant Very Relevant

How insightful is this suggestion?

Not Insightful Very Insightful

The screenshot displays a user interface for an opinion space. At the top, it shows the title 'Opinion Space' and the Silverman Research logo, along with 'Help' and 'Sign Out' links. Below this are three navigation buttons: 'Show more suggestions', 'Show Top Suggestions', and 'Show who I've rated'. On the right side, it indicates 'Suggestions Rated: 5'. The main content area features a participant profile for 'participant1855' with a close button. Underneath is a 'View Opinions' button and a discussion question: 'How could Unilever improve its IA policies?'. A text box explains that participants write their responses here to provide suggestions for improvement. There is a 'Flag for review' link. Below the text box are two rating scales: 'How relevant is this suggestion to you?' and 'How insightful is this suggestion?'. Each scale has a slider with 'Not Relevant' and 'Very Relevant' (or 'Insightful') labels. The background of the interface is dark blue with a visualization of suggestions represented as white and cyan dots of varying sizes, some of which are circled.

Participate Today!

Submit your comments on proposed regulations and related documents published by the U.S. Federal government. You can also use this site to search and review original regulatory documents as well as comments submitted by others.

Help improve Federal regulations by **submitting your comments**.

Are you new to the site?

Click the links below to get started.

- > [How do I find a rule?](#)
- > [How do I submit a comment?](#)
- > [How do I find my comment?](#)
- > [Do my comments make a difference?](#)

[Site Enhancements](#)

SEARCH for: Rules, Comments, Adjudications or Supporting Documents:

DiscoverText "the new geometry of text"

[our solution](#) | [industries](#) | [customers](#) | [data](#) | [pricing](#) | [news](#)

Try free for 14 days
1-800-936-0534

[sign up is free](#) [watch tutorials](#)

A Commenter's Checklist



[View Tips for More Effective Commenting](#)

Regulations With Comments Due Soon

- Today (79)
- Next 3 Days (113)
- Next 7 Days (192)
- Next 15 Days (427)
- Next 30 Days (787)
- Next 90 Days (1,102)

Newly Posted Regulations

- Today (97)
- Last 3 Days (98)
- Last 7 Days (556)
- Last 15 Days (1,0
- Last 30 Days (2,0
- Last 90 Days (6,1

Government Solutions

An invaluable tool for state and federal agencies to handle large numbers of documents.

Our electronic discovery tools help agencies to respond to legal exigencies. DiscoverText delivers key features for eDiscovery, FOIA and public comment response at a lower cost.

Identify data from multiple sources

- Federal Docket Management System Archives
- Import Plain text, Word, Excel, HTML, and PDF
- Emails from Exchange servers and Outlook files
- Public comments from YouTube and Ideascale
- News feeds and comments
- Facebook & Twitter feeds
- Blog and wiki content
- XML documents
- RSS feeds
- Instant messaging archives
- Microsoft SharePoint and IBM Lotus Notes
- Accounting databases

Ideal for FOIA processing

- Redact** sensitive, confidential or classified information
- Improve** communication between constituents and legislators
- Annotate** shared memos during document classification
- Auto-highlight** unique & offensive language
- Detect** threatening and other problematic communication
- Prevent spoliation**; all metadata is stored and searchable
- Store data** in raw and "native" formats
- Read** receipts generated on all documents
- Review** material as printed paper, a "native file" PDF format
- Review** single or multi-page TIFF images
- Adjudicate** documents or units of text coded by others
- Share projects** to create large private analysis teams

No software to install as our cloud-based application is securely hosted for you

Try **DiscoverText Free for 14 days**
No obligation and no credit card required

[Sign up](#)



Texifter Customers



[View our testimonials](#)

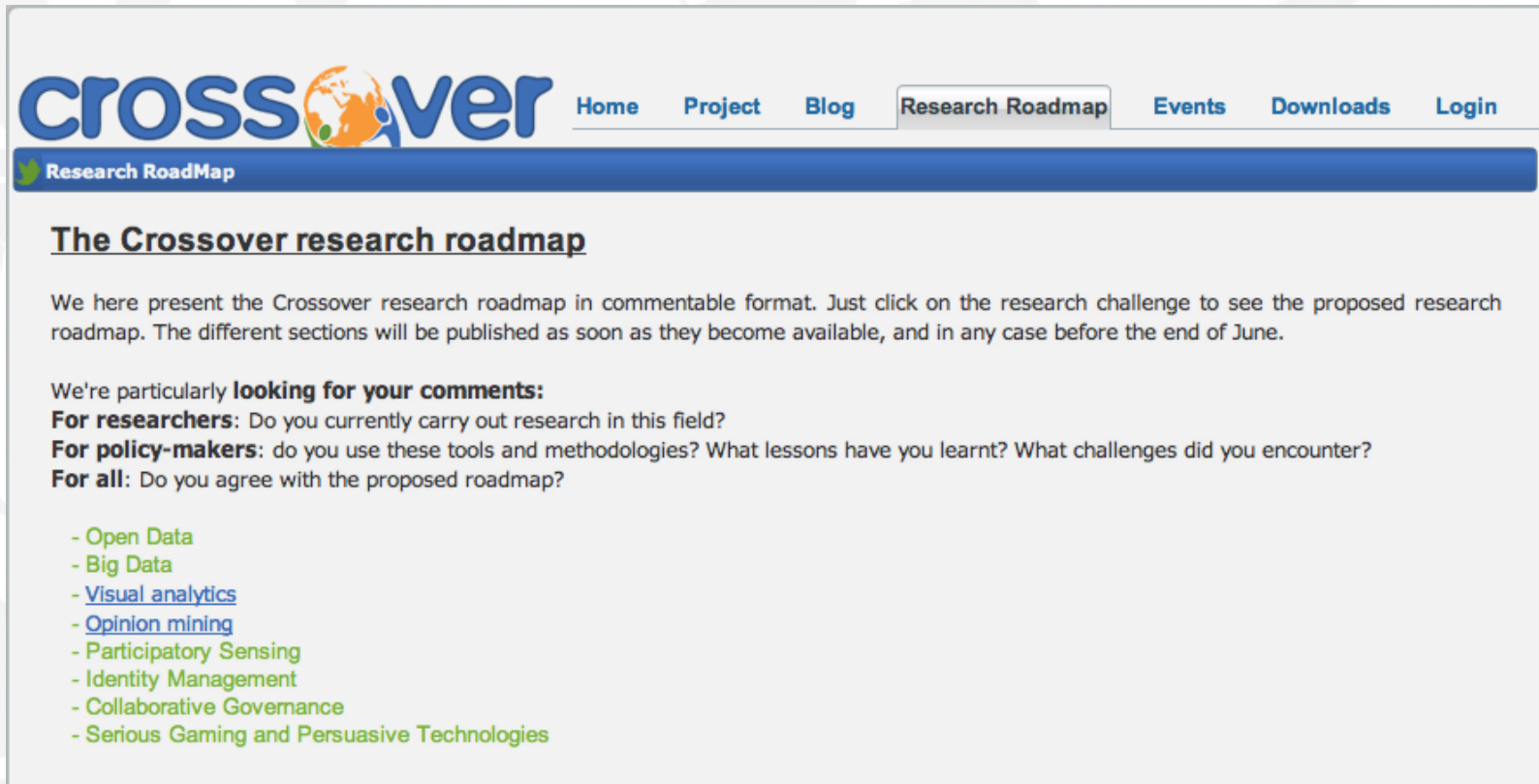
FUTURE CHALLENGES

- Reduction of human effort
- The detection of spam and fake reviews,
- The limits of collaborative filtering, which tends to identify most popular concepts and to overlook most innovative / out of the box thinking
- The risk of a filter bubble (pariser 2011)
- The asymmetry in availability of opinion mining software, which can currently be afforded only by organisations and government, but not by citizens. In other words, government have the means today to monitor public opinion in ways that are not available to the average citizens. While content production and publication has democratized, content analysis has not.
- The integration of opinion with behaviour and implicit data, in order to validate and provide further analysis into the data beyond opinion expressed
- The continuous need for better usability and user-friendliness of the tools, which are currently usable mainly by data analysts

Current free tools	Top market tools	Current research	Short term future research	Long term future research
<p>filtering opinion based on rating; assessing sentiments based on keywords; visual word counting Argument mapping and VAA</p>	<p>Machine learning + human analysis</p>	<p>Statistical + Semantic analysis through lexicon/corpus of words with known sentiment for sentiment classification Identification of policy opinionated material to be analysed Computer-generated reference corpuses in political/governance field Visual mapping of bipolar opinion Identification of highly rated experts</p>	<p>Visual representation Audiovisual opinion mining Real-time opinion mining Machine learning algorithms Natural language interfaces SNA applied to opinion and expertise Bipolar assessment of opinions Multilingual reference corpora Recommendation algorithms</p>	<p>Multilingual audiovisual opinion mining Usable, peer-to-peer opinion mining tools for citizens Non-bipolar assessment of opinion Automatic irony detection</p>

COMMENT THE ROADMAP

<http://www.crossover-project.eu/ResearchRoadmap.aspx>



The screenshot shows the 'Research Roadmap' page of the Crossover project website. The page features a navigation menu with links for Home, Project, Blog, Research Roadmap (which is highlighted), Events, Downloads, and Login. Below the navigation is a blue header with the text 'Research RoadMap'. The main content area is titled 'The Crossover research roadmap' and contains a paragraph explaining that the roadmap is in a commentable format and will be published by June. It also includes a section for 'looking for your comments' with specific prompts for researchers, policy-makers, and the general public. A list of research topics is provided at the bottom.

crossover Home Project Blog **Research Roadmap** Events Downloads Login

Research RoadMap

The Crossover research roadmap

We here present the Crossover research roadmap in commentable format. Just click on the research challenge to see the proposed research roadmap. The different sections will be published as soon as they become available, and in any case before the end of June.

We're particularly **looking for your comments:**

For researchers: Do you currently carry out research in this field?

For policy-makers: do you use these tools and methodologies? What lessons have you learnt? What challenges did you encounter?

For all: Do you agree with the proposed roadmap?

- Open Data
- Big Data
- [Visual analytics](#)
- [Opinion mining](#)
- Participatory Sensing
- Identity Management
- Collaborative Governance
- Serious Gaming and Persuasive Technologies