Project Acronym:	**CATALYST**
Project Full Title:	**Collective Applied Intelligence and Analytics for Social Innovation**
Grant Agreement:	**6611188**
Project Duration:	**24 months (Oct. 2013 - Sept. 2015)**

D3.5 Deliberation analytics

Deliverable Status:	**Final**
File Name:	**CATALYST_ D3.5.pdf**
Due Date:	**May 2014 (M8)**
Submission Date:	**June 2014 (M9)**
Dissemination Level:	**Public**
Task Leader:	**University of Zurich**

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

The CATALYST project consortium is composed of:

SO	Sigma Orionis	France
I4P	Imagination for People	France
OU	The Open University	United Kingdom
UZH	University of Zurich	Switzerland
EN	Euclid Network	United Kingdom
CSCP	Collaborating Centre on Sustainable Consumption and Production	Germany
Purpose	Purpose Europe	United Kingdom
Wikitalia	Wikitalia	Italy

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Revision Control

Version	Author	Date	Status
0.1	Mark Klein (UZH)	June, 2014	Draft
1.0	Stéphanie Albiéro (Sigma)	June 12, 2014	Quality check and submission to the EC

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Table of Contents

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Executive summary

The present document is a deliverable of the CATALYST project, funded by the European Commission's Directorate-General for Communications Networks, Content & Technology (DG CONNECT), under its 7th EU Framework Programme for Research and Technological Development (FP7).

This deliverable reports on Task 3.5, whose goal was to implement new forms of analytics to identify meaningful patterns in online deliberation, and map these patterns to personalised attention-mediation recommendations for the deliberation participants. In the following, the report will describe the technologies we have developed to achieve this goal, including:

- The metrics that UZH has implemented for finding meaningful patterns in deliberation activity
- A pattern-matching language for expressing when attention mediation actions should be taken

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

1. Deliberation Metrics

Our team developed a web-based metrics server that accepts requests consisting of:

- A representation of the deliberation map to analyse, including a record of the events (creating, editing, viewing, rating) performed by the deliberation participants on that map. This map is represented using the syntax in deliverable 3.1
- A specification of which metrics should be calculated for this map

More specifically, users should send use the HTTP protocol to POST a metrics request to the ACCEPT method on the metrics server. The REQUEST attribute of the POST should include a JSON object with the map you want to analyse plus the metrics you want for that map. The latter should be included as the value of the "@request" attribute of the JSON object (i.e. at top level alongside the "@context" and "@graph" attributes). The JSON syntax for the requests is:

"@requests" : [**requestlist**]

where

requestlist := request
request, requestlist

The response from the metrics server will be a JSON object giving you the metrics values you requested:

{"results": [**responselist**] }

where

responselist := response
response, responselist

The metrics descriptions, as well as the syntax for metrics requests and results, is given in the table below:

Request	**Response**	**Explanation**

{ "metric" : "post_space" }	{
	"metric" : post_space",
	"root" : "**post ID**"
	"data" : [[**postID, numlist**]*]
	}
		This gives the topic space coordinates for all the posts in the argmap branch. People who are interested in one post tend to also be interested in other posts with similar coordinates.

{ "metric" : "user_space" }	{
	"metric" : "user_space",
	"data" : [[**userID, numlist**]*]
	}
		This gives the topic space coordinates of the topic user. Users with similar coordinates have similar interests.

{ "metric" : "expertise",
"user" : "**user ID**" }
	{
	"metric" : "expertise",
	"user" : "**user id**",
	"data" : {
		Specifies the average rating for the posts a user has contributed in a given topic. 0 means the user did not

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

	"topic1" : **number,** "topicN" : **number** } }	contribute anything to that topic.

{ "metric" : "similar",
"user" : [**user ID**] }
	{
	"metric" : "similar",
	"user" : **user ID**,
	"data" : [**ulist**]
	}
		Finds users who are interested in (active with) the same topics as the given user.

{ "metric" : "recommendations",
"user" : "**user ID**" }
	{
	"metric" : "recommendations",
	"user" : "**user id**",
	"data" : {
	"**post id**" : **number,**
	"**post id**" : **number**
	}
	}
		Recommendations for posts that may interest a given use (based on what he/she looked at in the past, as well as on what has been looked at by people with similar interests). One recommendation per line, with a number that gives the "strength" of the recommendation from 1 (weakest) to 10 (strongest).

{ "metric" : "controversy",
"issue" : "**issue id**" }
	{
	"metric" : "controversy",
	"issue" : "**issue id**",
	"data" : **number**
	}
		A score that specifies how controversial the discussion for an issue is, ranging from 0 (low controversy) to 1 (highly controversial).

{ "metric" : "maturity",
"issue" : "**issue ID**" }
	{
	"metric" : "maturity",
	"issue" : "**issue id**",
	"data" : **number**
	}
		A score that specifies how mature the discussion for an issue is, ranging from 0 (not at all mature) to 1 (highly mature).

{ "metric" : "narrowing",
"issue" : "**issue ID**" }
	{
	"metric" : narrowing",
	"issue" : "**issue id**",
	"data" : **number**
	}
		A score that specifies to what extent the discussion for an issue has focused on a single idea at the expense of the others (0 = low groupthink score, 1 = high). Calculated using a gini coefficient approach.

{ "metric" : "bias_space",
"root" : "**post ID**" }
	{
	"metric" : "bias_space",
	"root" : "**post ID**" ,
	"data" : {
	"bias1" : [**plist**],
	"biasN" : [**plist**]
	}
	}
		This defines the set of "biases" for a given branch in a discussion, where each bias is a set of posts whose ratings tend to be correlated. For example, people who like solar may also tend to like wind power, representing a bias with respect to renewable energy. The first biases in the list have more statistical significance than those lower down. If the branch root is not given, this metric looks at the entire discussion.

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

{ "metric" : "bias",
"user" : "**user ID**" }
	{
	"metric" : "bias",
	"user" : "**user id**",
	"data" : {
	"bias1" : **number,**
	"biasN" : **number**
	}
	}
		This gives a user's position on each bias dimension. Users with similar coordinates have similar biases. A highly balkanized or polarised discussion will show up as a scatter plot, of user biases, where people fall into distinct clumps. We could convert this scatter plot into a metric that captures the degree of balkanization by measuring the degree of clustering in the plot.

{ "metric" : "agreement",
"root" : "**issue ID**",
"users" : [**ulist**] }
	{
	"metric" : "agreement",
	"issue" : "**issue id**",
	"users" : [**ulist**],
	"data" : **array**
	}
		A table, for each issue, where the rows and columns represent users, and the cells represent how much each user pair agrees about the best ideas for the issue. This could be visualised as a force-directed graph where nodes = users, where agree and disagree links have different colors, and where users that agree with each other are placed close to each other and far from those they disagree with. Balkanization and polarisation would show up as strong clustering in the graph.

{ "metric" : "grounded",
"idea" : "**idea ID**",
"users" : [**ulist**] }
	{
	"metric" : "grounded",
	"idea" : "**idea id**",
	"users" : [**ulist**],
	"data" : **number**
	}
		A score that measures to what extent an idea's average rating is consistent with the ratings for the underlying arguments. This can be done for the ratings from an individual user, for a group of users, or for all users. If no user IDs are specified, it does the calculation for all users. (0 = inconsistent, 1 = consistent)

Where:

plist := **post ID**
post ID, plist

ulist := **user ID**
user ID, ulist

array := [**vectorlist**]

vectorlist := **vector**
vector, vectorlist

vector := [**numlist]**

numlist := **number**
number, numlist

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

These metrics are calculated using several key technological building blocks, including:

- *Singular value decomposition* (SVD): SVD [1] is a matrix analysis technique that can compress a hyper-dimensional data set into one with a smaller set of dimensions. We use it create a low dimensional "topic space" capturing which posts attract the attention of which users. With this, we can readily do such things as identify which users have similar interests, as well as which posts are strongly related. We use the SVD code from the well-respected (and widely-used) GNU Scientific Library.
- *Hierarchical cluster analysis:* This technique is used to identify *clusters* in the "topic space" identified by SVD, so we can develop a better understanding of the user communities sub-structure.
- *Pattern-matching*: we developed a pattern matching language (described below) to look for patterns in participant activity in multi-user argumentation maps.
- *Bayesian belief propagation:* This technique [3] can be used to fuse and propagate the impact of new evidence and beliefs through logical networks so that each proposition eventually is assigned a certainty measure consistent with the axioms of probability theory. This can be used to infer what participant ratings on ideas and arguments imply about the degree of support for the arguments and ideas above them in the argument map.
- *Variance statistics*: we use such well-known statistical tools as T-tests to identify which patterns represent statistically significant, as opposed to potentially just random, activity patterns.

2. Deliberation Pattern-Matching Language

We developed a highly expressive query language specifically designed for finding patterns in user activity in multi-user argument maps, building on previous work by team members [2]. The language includes query clauses, in particular, that can match users, posts, links between posts, and events, and can also include Booleans as well as arbitrary Common Lisp predicates. The syntax for this language is:

```
query :=  ( - <clause> - )

clause := (post <var> isa (<type>*))
          (user <var> [isa (<role>*)])
          (link <var> <var> [<recur>])
          (event <var> [post <var>] [who <var>] [what <op>])
          (mentions <var> <string>)
          (when <lisp code>)
          (thereis <query>)
          (never <query>)
          (and -<query>-)
          (or -<query>-)
          (let <var> be <lisp code>)
          (let <var> {find-all find-one} <query>)

type   := {issue idea pro con comment}

role   := reader | rater | commentator | author | senior-author | moderator | sysadmin

recur  := * | +
```


The CATALYST project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

op := ({ viewed created edited commented rated }*)

result := (- <bdg> -)

bdg := (<var> . <value>)

var := ?<symbol>

Using this language, it is easy to specify queries for events that require user attention.

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

References

[1] Golub, G. H., & Van Loan, C. F. (2012). *Matrix computations* (3). JHU Press.
[2] Klein, M., & Bernstein, A. (2004). Towards High-Precision Service Retrieval. *IEEE Internet Computing Journal*, *8(1)*(1), 30-36.
[3] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann.

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188