



Mastering Data-Intensive Collaboration and Decision Making

FP7 - Information and Communication Technologies

Grant Agreement no: 257184
Collaborative Project

Project start: 1 September 2010, Duration: 36 months

D6.4.1 – Report from the evaluation of use case #3 (first version)

Due date of deliverable: 29 February 2012
Actual submission date: 20 March 2012
Responsible Partner: PUB
Contributing Partners: NEO, FHG, BRF, CTI, UPM

Nature: Report Prototype Demonstrator Other

Dissemination Level:

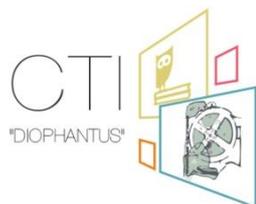
- PU : Public
- PP : Restricted to other programme participants (including the Commission Services)
- RE : Restricted to a group specified by the consortium (including the Commission Services)
- CO : Confidential, only for members of the consortium (including the Commission Services)

Keyword List: Evaluation, social media monitoring, sentiment analysis, unstructured web 2.0 data, data management, collaboration, decision making



The Dicode project (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation programme (ICT/SO 4.3: Intelligent Information Management).

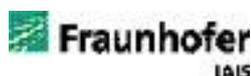
The Dicode Consortium



Computer Technology Institute & Press "Diophantus"
(CTI) (coordinator), Greece



University of Leeds (UOL), UK



Fraunhofer-Gesellschaft zur Foerderung der angewandten
Forschung e.V. (FHG), Germany



Universidad Politecnica de Madrid (UPM), Spain



Neofonie GmbH (NEO), Germany



Image Analysis Limited (IMA), UK



Biomedical Research Foundation, Academy of Athens
(BRF), Greece



Publicis Frankfurt Zweigniederlassung der PWW GmbH
(PUB), Germany

PUBLICIS

Document history			
Version	Date	Status	Modifications made by
1	08-02-2012	First draft	Ralf Löffler, PUB
2	20-02-2012	Second draft, evaluation results incorporated	Anastasia Kastania, BRF Georgia Tsiliki, BRF Sophia Kossida, BRF Ralf Löffler, PUB
3	29-02-2012	Sent to internal reviewers	Max Jakob, NEO Doris Maassen, NEO Nikos Karacapilidis, CTI Anastasia Kastania, BRF
4	14-03-2012	Internal reviewers' comments incorporated, sent to SC	Ralf Löffler, PUB
5	20-03-2012	Final version (approved by SC, sent to the Project Officer)	Ralf Löffler, PUB

Deliverable manager

- Ralf Löffler, PUB

List of Contributors

- Ralf Löffler, PUB
- Anastasia Kastania, BRF ((i) Dicode services user trials organization, (ii) Dicode services user trials implementation in Greece, (iii) Redesign of questionnaires of D6.1 according to EC Reviewers comments, (iv) Design of the Dicode user trials protocol, (v) Construction of the Dicode evaluators manual, (vi) Production of the descriptive statistics evaluation results (statistics for deliverable body and appendices both for UC1 & UC3), (vii) Implementation of the reliability analysis (both for UC1 & UC3))
- Georgia Tsiliki, Sophia Kossida, BRF
- Nikos Karacapilidis, Manolis Tzagarakis, CTI
- Guillermo de la Calle, Miguel García-Remesal, UPM
- Stefan Rüping, FHG
- Doris Maassen, NEO

List of Evaluators

- Max Jakob & Doris Maassen, NEO
- Nikos Karacapilidis, CTI
- Anastasia Kastania, BRF

Summary

This deliverable reports on the first round of the evaluation of Dicode services for Use Case 3: Opinion Mining from Unstructured Web 2.0 Data (OMUWD). The evaluation process was performed by using properly formulated metrics and instruments, which have been described in D6.1 and follow the specifications of D2.2. The deliverable also includes an updated description of Use Case 3. The assessment of the already developed Dicode services raised interesting technical and methodological issues which are also described in this deliverable.

Table of Contents

1	Introduction.....	5
1.1	Context	5
1.2	Objectives.....	5
1.3	Structure.....	5
2	Overview of Dicode Use Case #3.....	6
2.1	Description of Dicode Use case #3	7
2.2	Input data.....	8
2.3	Output data.....	9
2.4	Dicode’s implemented services	9
3	Evaluation process and feedback.....	11
3.1	Interviewing ten marketing professionals	11
3.1.1	Interview design	11
3.1.2	Interview results	12
3.1.3	Implications for the Dicode approach	14
3.2	Detailed evaluation of Dicode services.....	15
3.2.1	Evaluation process.....	15
3.2.2	Detailed evaluation feedback.....	15
3.2.2.1	<i>Blog pre-processing service</i>	15
3.2.2.2	<i>Topics service</i>	20
3.2.2.3	<i>Keytrends service</i>	24
3.2.2.4	<i>Twitter pre-processing service</i>	28
3.2.3	Reliability analysis.....	32
4	Future work directions	33
	References.....	34
	Appendix: Evaluator’s Guide.....	35

1 Introduction

1.1 Context

This deliverable reports on the evaluation of the initial version of the Dicode services for the needs of Use Case #3: “Opinion Mining from Unstructured Web 2.0 Data” (OMUWD), as these have been developed in the context of WP2-WP5. This is the first of two deliverables reporting on the validation and assessment of the Dicode services for the needs of OMUWD. While the focus of this deliverable (D6.4.1) is on the first round of the evaluation of the Dicode services, deliverable D6.4.2 (due in month 32) will present the evaluation of their enhanced version and also comment on how the feedback provided in the first evaluation round was taken into account. The project’s evaluation activities will be summarized in deliverable D6.5 (due in month 36), together with an assessment of the project from external experts and bodies.

1.2 Objectives

The purpose of this document is to present the results of the first evaluation round of the developed Dicode services related to the OMUWD use case (Task 6.5), in order to assure their usability and accessibility. The cost-effectiveness and readiness of Dicode services for the market are also assessed. Through the OMUWD use case, we aim to validate the Dicode services as far as the automatic analysis of voluminous amount of unstructured information is concerned. Data for this case are primarily obtained from crawling the Web (blogs, forums, and news). In addition, different APIs from various Web 2.0 platforms, such as micro-blogging platforms (Twitter), and social network platforms (Facebook) are used. An updated description of the OMUWD use case is also presented in this document to reflect the current developments in the marketing community. The Dicode services are evaluated based on how they can improve the processing of OMUWD in collaboration and decision making settings, so as to support marketing decisions and strategies.

The evaluation of the Dicode services is performed based on a series of Key Success Indicators (KSIs), already reported in D6.1, overall aiming to assess:

- maturity of the technology used through Technology Readiness Level (TRL);
- dissemination and exploitation activities;
- usability and acceptability of Dicode services in the OMUWD use case.

In this deliverable, emphasis is given to those KSIs measuring usability and acceptability of the Dicode services in the related marketing community. The evaluation feedback collected and critically presented in this deliverable will serve as a guideline for the service providers to update the Dicode suite of services and specifications as well as to improve their quality.

1.3 Structure

Section 2 starts with an updated description of OMUWD. Section 3 provides information about the overall evaluation process of OMUWD. It comprises two parts, one summarizing the opinions of ten marketing professionals on the overall approach followed in the Dicode project with respect to the OMUWD use case, and one reporting on the evaluation of the specific Dicode services for this case. Section 4 reflects on the feedback of the evaluation procedure and outlines directions for the improvement of the related Dicode services.

2 Overview of Dicode Use Case #3

Web 2.0 is heavily challenging existing marketing and communication paradigms. For a long time, brands communicated in a top-down mode: firms have been the sender, while consumers were the receiver or addressees. (Figure 2.1 - left part). These brands controlled the conversation in terms of topics, critics and media penetration. Today, such a model of communication is obsolete; instead a cross-linked communication model is adopted (Figure 2.1 - right part). Accordingly, brand managers from agencies and companies have to reassess and realign their tools in order to cope with the rise of the Web 2.0 era.

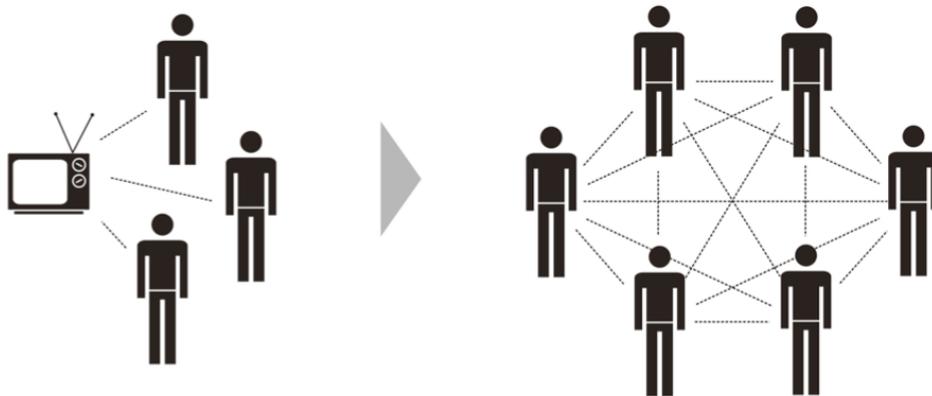


Figure 2.1: Radical Change in Communications: From the sender-receiver principle to cross-linked communication (Löffler/Wittern in *Markendifferenzierung*, Völckner, F./Willers, C./Weber, T., Gabler Verlag 2010)

The fusion of World Wide Web and mobile phones changed people's communication behaviour and habits. New opportunities made them even more permissive, demanding and self-confident: they communicate always and everywhere, built networks, spread their opinion, generate their own content and continuously demand news and interaction. Consumers become confident. They actively influence the perception of the brand: by now, more than 25% of the Top 20 brands online content is user generated (Social Media Research Wave 3, Universal McCann, March 2008 - see <http://www.slideshare.net/mickstravellin/universal-mccann-international-social-media-research-wave-3>). In addition the consumers' voice is much more reliable than that of the company. In other words: the consumers have taken over the power. This forces companies to invest in additional resources with the aim to understand the "whom", "what" and "where" in the social web.

Through Social Media Monitoring, marketing and strategic planning have a complex and versatile tool for listening to the Internet and the consumer's voice. All comments related to brands and companies, reviews of products or conversations between consumers about brands can be located and filtered by utilizing appropriate software tools. For this purpose, a variety of services with different priorities and qualities have been developed in recent years. Some are freely available (e.g., Google Blog Search or Technorati), while for others a fee is required (e.g., Radian&, Sysomos and SM2). Thus, Social Media Monitoring tools seem to be the right solution to rebuild a company's capacity to listen to and interact with consumers.

A big worldwide survey from the Meltwater Group (involving 450 Companies) shows that more than 84% of the participating companies would like to invest in monitoring such data, but less than 20% actually do (Future of Content, The Meltwater Group, November 2010 - see http://meltwaterproducts.com/reports/Meltwater_Future_of_Content_Report.pdf). Two reasons appear to be the strongest barriers:

- the quality of the existing solutions is not sufficiently reliable to build a solid basis for marketing and business strategies, and
- integration of Social Media Monitoring into existing company structures and marketing procedures is often difficult.

2.1 Description of Dicode Use case #3

While many experts are complaining about the missing reliability of data due to needed improvement in filtering, sentiment analysis and influencer analysis, there is another important issue that has to be addressed: *the need for a working environment that successfully involves different parties into the social media monitoring process*. The goal of Use Case 3 is to create a collaborative environment and a set of services for the automatic analysis of a voluminous amount of unstructured information. Such a tool will allow decision makers to make better, more informed and quicker decisions, by displaying the aggregated information according to their needs.

In the following paragraphs, we describe a typical decision making process in social media marketing through two scenarios concerning the current and the desired way of working (after implementing the Dicode solution). In these scenarios, three parties - representing three typical key functions in a Social Media Monitoring process - are involved:

- Sarah, Marketing Department;
- Frank, Social Media Analyst, and
- Jan, Social Media Engager.

All three are very committed and show an impressive track record in their specific area - but results haven't been satisfying so far: Sarah argues that her Marketing Department still hasn't felt an actual benefit from Social Media Monitoring and that she's not in the loop about the current actions, while Frank is complaining about the missing consistency of the data and the vague results of the influencer analysis. Jan, however, is blaming Frank for the lack of (consumer) insights that he can build on in his Social Media activities. Reasons for this disappointing experience are:

- The sequential way of working, which causes huge delays;
- There is almost no communication in between the three parties and thus too little collaboration;
- The work often depends on a single person - there is no knowledge transfer or appropriate infrastructure that ensures knowledge backup;
- There is only little interaction with consumers - respectively, there is no transformation of acquired data results to valuable insights;
- No reliable results due to missing data quality.

All three parties basically need to work closer together in every stage of a typical project: listening to the consumers, identifying risks and opportunities, creating solutions, making decisions about communication and engaging with the social web. Only a working infrastructure and an appropriate suite of services for close collaboration can lead to such a structured process and thus to actionable results (Figure 2.2).

2.3 Output data

The data must be filtered, cleansed, enriched, and analyzed depending on the type of data source. Different analysis modules must be combined in a way that allows each module to work with and process the metadata and other information obtained from other modules. This requires that each module is designed in such a way that it has a global system interface to share metadata with the other modules. After passing through the different modules, the data must be aggregated, summarized and displayed in such a way that the user can make sense of it all. The system should provide a knowledge base obtained from the aggregated and summarized data. The system should provide a new level of search capabilities (navigation in aggregated and summarized data). This will allow for more well-informed and quicker decisions to be made. Finally, once data has been aggregated and summarized, different visualizations are needed to meet the needs of different end users.

2.4 Dicode's implemented services

As far as the OMUWD use case is concerned, online communication is direct, unvarnished and honest; every single voice can be tracked and be heard, something that offers a big source for enlightening insights and marketing opportunities. On the other hand, the usability of the underlying tool and the possibility to provide easy access and make sense of the information delivered are key issues for the associated end user group (which are not IT professionals). The challenge of the Dicode project is to deliver services that cover these demands. Dicode services developed so far for the OMUWD use case are:

- **Workbench service** – a web application built for all Dicode use cases, which provides a common graphical interface to access and use heterogeneous services. The Dicode workbench is the integration platform for all Dicode data analysis and collaboration services (D4.1.1, D4.2.1, D5.4.1). Within the Workbench the **Storage service** is included, an integrated service built for all Dicode use cases to comfort the sharing and exchange of information (files, reports, etc.) in data-intensive and cognitively-complex settings. The service provides all functionalities needed to allow permanent and reliable storing of files as well as their accessibility (D5.4.1).
- **Collaboration & Decision Making Support services** – integrated services built for all Dicode use cases. The **Collaboration Support** service exploits the reasoning abilities of humans to facilitate sense-making of the Dicode data mining services results and capitalize on their outcomes (D4.1.1, D4.2.1). The **Decision Making Support** service turns information and knowledge, currently only for a limited set of fixed discourse moves and message types, machine interpretable in order to allow active participation of the system in collaborative activities (D4.1.1, D4.2.1).
- **Forum Summarization service** – a standalone service built for the needs of Use Cases #1 and #3. The service takes a cluster of discussion threads as input from the Workbench service and identifies the most prominent terms (topics) in the discussion threads that belong to the cluster (D2.1, D4.2.1). The identified topics can be used by the users to identify what the main theme and the discussion threads in that cluster are about.
- **SubGroup Discovery service** – a standalone service built for the needs of Use Cases #1 and #3. The service searches for subgroups in any user provided data by searching the rules that cover many target value examples and few non-target value examples. The service supports two different Subgroup Discovery algorithms (D3.2.1, D4.1.1, D4.2.1).

- **Recommendation service** - a standalone service built for all Dicode use cases. The service recommends similar users or documents from log file data (D3.2.1, D4.1.1, D4.2.1) based on similarity models learned by using the Dicode Similarity Learning Service.
- **Blog-preprocessing service** - It returns a condensed representation of weblog entries containing only significant nouns. Optionally, one can query for a search term (a regular expression). Additionally, one can filter the results by language, start- and end-date.
- **Topics service** - It is intended to give the user a quick albeit superficial overview of the thematic content of a document collection. Large document collections of twitter messages and/or collections of consumer discussions in Internet blogs are scanned for the thematic content that is interesting for the analyst. Each of the detected topics is described by a list of keywords sorted by keyword importance (i.e. probability of occurrence) for this topic.
- **Keytrends service** - The Keytrends service returns metadata about tweets on a selected day. The following metadata is currently available: (a) Hashtags (Top hashtags), (b) Language (Languages of tweets), (c) Country (Country code of Twitter user), (d) Place (Places of Twitter user [only available for few tweets]) and (e) Urls (Urls mentioned in tweets). In each case, the top 100 (or less) values are returned in descending order.
- **Twitter preprocessing service** - It returns a condensed representation of Tweets containing only significant nouns. Optionally, one can query for a search term (a regular expression). Additionally, one can filter the results by language, start- and end-date. During development, the full tweet is also returned. Starting from the evaluation, the Tweet text will be removed from the result.

Evaluation results for the first five services above are presented in D6.2.1 (since these services concern both Use Case #1 and Use Case #3, these results are not repeated here). Evaluation results for the last four services are reported in Section 3 below.

3 Evaluation process and feedback

Parallel to the user trials conducted to evaluate the current implementation of the Dicode services for all three use cases, a series of interviews with marketing professionals has been carried out to re-assess the overall Dicode approach for the needs of OMUWD case. Specifically, to make sure that this case's objectives are being met, ten expert interviews took place in January and February 2012. In addition to the general assessment of the future of social media in the context of the project, the aim was to analyze the requirements and the specific challenges of social media monitoring. In these interviews, the overall Dicode approach was presented and the proposed Dicode workbench solution was critically analyzed. The design and results obtained from these interviews appear in Section 3.1. Section 3.2 reports on the evaluation of the last four services mentioned in Section 2.4.

3.1 Interviewing ten marketing professionals

3.1.1 Interview design

The overall objective of these interviews was to ensure the usability of the Dicode approach for the Social Media Monitoring process. Our specific target group consisted of ten high level marketing professionals. Understanding their requirements for this process is certainly important for the successful exploitation of Dicode services. Usually, such people have limited time to spend and are not used to work towards evaluating innovations in early stages. For them, the "big picture" is by far more important than details. We carried out this experiment through individual face-to-face and telephone interviews in order to have the possibility to clearly explain our approach, answer upcoming questions, and obtain a deep understanding of their needs and judgements. In marketing, it is customary to review the benefits of innovations based on the main concept, as long as the final, perfectly functional and user-friendly implementation does not yet exist. In other words, it was the "big picture" and not specific services that were evaluated through this experiment.

Interviews' participants are in leadership positions in communication agencies and industry with huge experience in the field of digital communications. Details about their roles and associated companies are summarized in Table 3.1.

Name	Role	Company	Description
Clemens Sommer	Managing Director	Creative Analytic 3000	a company for consumer research, internationally active, about 30 employees, clients: Deutsche Bank, Procter & Gamble, Deutsche Telekom, Novartis, Bongrain etc.
Tanja Grubner	Senior Brand Manager DACH	SCA Group	develops, produces and markets personal care products, tissue, packaging, publication papers and solid-wood products, and has sales in more than 100 countries with 45,000 employees worldwide
Daniel Putsche	Managing Director Germany	Mags	Supports advertising agencies in terms of digital communications, with more than 60 employees, headquarter in Sofia
Dieter Romatka	Chairman	MEC Access Germany	Media Agency that employs 4.400 people in more than 80 countries, Clients: Campbell, Microsoft, Honda, Sony, Paramount etc.

Oliver Jeschonnek	Founder & CEO	Socialect	Start up, consulting agencies and industry due to Social Media, Client: IngDiBa
Prof. Dr. Benjamin Schwenn	Managing Partner	ISK - Institut für Strategie & Kommunikation	A strategy consultancy with focus on brands, communications and change management, Clients: Wepa, TraffiQ, Cofresco etc.
Achim Ewers zum Rode	Manager Brand Strategy & CI Europe	Opel AG	Car manufacturer, part of General Motors, Opel has about 25.000 employees in Germany
Anja Merkelbach	Media Director	Jaron GmbH	full service online marketing agency, about 60 employees in Berlin and Frankfurt, Clients: Audi, Allianz, Ergo, DPD etc.
Michael Kurz	Senior Strategic Planner	Syzygy AG	Online Marketing Agency with about 300 employees in Germany, Poland and UK, Clients: Mercedes-Benz, Commerzbank, AVIS, O2, Continental etc.
H. P. Simon	CEO	Pixelgenau AG	Innovative Agency with focus on online, mobile and social media, about 15 employees, Clients: Subaru, Deutsche Telekom, SHR

Table 3.1. Marketing professionals interviewed.

3.1.2 Interview results

This section reports on the issues raised and summarizes the results of the experts' interviews. The interviews were structured around four main questions.

(i) How would you rate the impact of social media on the communications industry?

There was consensus that social media have already changed the communications landscape significantly. This opinion was without exception shared by all participants. Companies must address these new challenges to cope with shift from a transmitter-receiver communication to a participatory communication landscape. This requires, above all, greater transparency of communication and supply, in order to build trust with the critical consumer. Companies need to change and learn to communicate with consumers on equal terms. A representative comment was: *"Social Media has already and will further change the way companies will act and communicate"* (Participant, Agency).

However, there are also voices that see social media as a "hype", sometimes resulting in a hasty urge to launch social media campaigns, without even considering objectives, measures and appropriate tools. Marketing has understood that there is a need to enter Social Media to promote brand and products, but still needs an adequate framework in terms of signing clear rules which give guidance of how and when to interact with customers. This seems to be the biggest future challenge for many of the respondents.

(ii) Does your company use Social Media Monitoring (SMM)? If yes, in what way? Quality of the data? If not, why not?

First of all, participants were familiar with the term "social media monitoring". For many of the respondents working in communication agencies, the monitoring is part of the portfolio of the company. Most of the respondents said that a good technical data analysis is

important; the core service, however, is the strategic analysis and the conclusions that can never be fully automated. For analysis, professional monitoring tools like Radian6, Sysomos and Vico are used. The quality of such tools is often questioned, particularly the reliability of many functions, such as sentiment analysis which is often seen as inadequate. A representative comment was: *"We are investing in Social Media Monitoring, but not in a way that it is beneficial for my department"* (Participant, Industry - Brand Communication).

Participants that are neither offering SMM as a service nor making regular use of SMM in their company, use self-made solutions and individual approaches. For this, they use freely available tools, which cover only a part of the spectrum of possibilities, but are mostly sufficient for simple tasks. The willingness to pay for professional monitoring tools is rather low. Others complained that their companies have invested in SMM but it is used for different purposes.

(iii) What are the biggest challenges for SMM in the future? Are there areas of the SMM process that need improvements?

Information value: Even participants with traditional market research background have difficulties to accept Social Media Monitoring as a representative research tool (*"Social Media Monitoring is not a representative research tool"* (Participant, Market Research)). These participants are highly trained in market research patterns, in that they look especially at weak points in information value and describe SMM as highly subjective or even arbitrary.

Data reliability: The respondents, who have worked intensively with SMM tool and their evaluation, can directly identify the (technical) weaknesses of the current monitoring tools. This especially concerns the sentiment analyses, which is crucial for the representation of user attitude towards the brand or brand related topics (*"Technical challenges are a crucial success factor for the future - I need to rely on the sentiment analysis"* (Participant, Agency)). Current solutions are not sufficiently reliable to conclude sustainable strategies and action steps. Language recognition and the recognition of irony, for example, are important points. Furthermore, the influencer's analysis is pointed out, which is somehow possible, but still not usable in a business context.

Usability: Usability is understood by participants in terms of a more practical and easier linking and presentation of data (*"It should be easier to display data in different context"* (Participant, Agency)). The opportunity to bring the filtered data in a meaningful context with the existing monitoring tools is really good. Different correlations and similar functions are unfortunately still limited. There is still a big need for development. The same applies to the possibility to connect the tool to other programs.

(iv) Suppose that you could use this tool (Dicode Workbench) in order to structure your SMM process in this way. What advantages do you see? Would there be any disadvantages?

The Dicode overall approach and the Dicode workbench were presented to the interviewees as a verbal concept. The services were described in a rather general way. The interviewees answered questions related to the Dicode approach. Emphasis was given to questions related to the benefits in terms of data reliability and collaboration. Generally speaking, the respondents thought that the Dicode project has focused on the right objectives in addressing the issue of data quality and the required collaborative working environment.

Data reliability was considered as a basic prerequisite for the successful placement of social media monitoring in the market. More precisely, sentiment analysis is considered to be the most important factor (*"A reliable sentiment analysis is the most challenging mission for software developers"* (Participant, Agency)). It is the core function and allows a quick insight into a particular brand or a brand-relevant topic. Marketing representatives think that the fragmentation of social media is one of the main challenges that should be addressed in the future. In particular, this refers to the problem of locating and extracting content and information that is really important for the brand. However, some respondents were sceptical about whether this improvement concerning data reliability can be effectively implemented.

The collaborative work platform in the sense of a virtual "war room" is picked by most to be very promising. In particular, it was revealed that the proposed Dicode platform works towards simplifying the process of coordination of planning, which is a very complex and data intensive issue. As a critical pre-condition, the participants named an interface which is simple to use and clearly structured. This is especially important to enable people to participate, who do not use SMM regularly. Some participants recommended a kind of "dashboard" that brings together all the relevant information and represents it clearly.

Moreover, the participants saw the Dicode approach's potential to improve efficiency and save resources and costs. Some of the participants, however, were sceptical when it comes to the integration of all involved parties for three reasons:

- not every detail is relevant for all parties;
- not every information should be transparent to all, and
- too much information can make the decision process slower.

Especially the marketing managers pointed out their need for brief and consolidated data as a basis for decision-making (*"I get around 300 emails a day, I'm attending meeting after meeting, I just don't have the time to familiarize with a complex and data intensive tool"* (Participant, Industry)). As a recommendation they proposed:

- a tool which offers different work environments, with different accesses for the involved parties, and
- the introduction of an additional open environment, which explains the conclusions of the Social Media Analysts in detail if wanted /needed. In this context, the description of the mind map view was highly appreciated (*"I need a quick update regarding the relevant topics, but I can't follow the conversation every day - a mind map with history would be great"* (Participant, Industry)).

3.1.3 Implications for the Dicode approach

The evaluation showed the big potential of the Dicode approach for the marketing industry; however, three areas for improvement were identified:

- For this target group, it is absolutely crucial that the relevant data is presented in a consolidated and user friendly way;
- The idea of collaboration and related services was seen very positively; the possibility of providing - at the same time - private areas would make it easier for the marketing professionals to adopt them;
- Last but not least, the capability of integrating the data with further applications e.g. Customer Relationship Management Systems (CRM) would be highly appreciated.

3.2. Detailed evaluation of Dicode services

3.2.1 Evaluation process

The Dicode user trials were conducted successfully in the four participating countries of the project using the Dicode Evaluation Framework. Use case partners and technical partners invited participants to join the user trials. Due to the fact that the majority of the up to now developed RESTful Dicode Services produce output in JSON format (<https://tools.ietf.org/html/rfc4627>; Crockford, 2006) the target population was defined to be computer scientists/informatics scientists with knowledge in Web Programming/Web Engineering. The sampling method used was Snowball sampling (Kitchenham & Pfleeger, 2002).

The evaluators group comprised 44 evaluators from Greece, 13 evaluators from Spain, 3 evaluators from Germany and 1 evaluator from the United Kingdom. To assure the fairness of the Dicode Services trials, the following strategy was applied:

- Ethical issues were applied in trials design with human participants (the right not to suffer harm, the right to full transparency, the right of self decision, the rights of privacy, anonymity and confidentiality).
- The three questionnaires appearing in D6.1 that were applied in the conducted survey research were adapted, including evaluator's confidence measurement in most of the qualitative questions and request for free-text comments on the improvement of the Dicode Services (see the Evaluators' Guide in the Appendix). Two additional questionnaires were designed and tested for the Forum Summarization Service.

The evaluation analysis results that are summarized in sections 3.2- 3.3 have been produced using the Statistical Package for the Social Sciences (SPSS 17v) routines.

3.2.2 Detailed evaluation feedback

This section includes summary statistics for all OMUWD case evaluations as well as representative qualitative feedback from the questionnaires. The tables and figures included summarize the quantitative evaluators' responses. Detailed statistics about the completed questionnaires have been uploaded to the Dicode wiki (<https://wiki.dicode-project.eu/display/DIC/WP6+Evaluation+results>).

When a specific service evaluation was considered, evaluators assessed the scenario provided, the acceptability and the overall quality of the service. Answers to the quantitative questions of the questionnaires are given in a 1-5 scale (Quality, Acceptability and Accessibility questionnaires), where 1 stands for 'I strongly disagree' and 5 for 'I strongly agree', and a 0-10 scale (Usability questionnaire), where 0 stands for 'none' and 10 for 'excellent' (Nielsen, 1991; Norman, 1998). The evaluation outcome is presented for each service separately. Note that missing data were not imputed.

3.2.2.1 Blog pre-processing service

OMUWD evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.1, 3.2 and 3.3, summarize the evaluators' responses relative to the quality, acceptability, accessibility & acceptability of the blog pre-processing service. In these figures, the median values of the evaluators responses are shown; the

original answers were given in a 1-5 scale (ranging from 1: strongly disagree to 5: strongly agree). For the quality of the service, the evaluators agree (median: 4, mode: 4) that the service addressed the data intensive decision making issues and are satisfied with the performance of the service. The evaluators are neutral about whether the objectives of the service are met, the service is novel to their knowledge, and they are overall satisfied with this service. In particular, related to the question of how satisfied evaluators were with performance of the service, comments such as *"In terms of time completion I am satisfied. But the results are not very understandable"* (GR1) summarize the evaluators feedback and explain the neutral score.

As long as the acceptability of the service is concerned, the evaluators are between disagree and neutral that the interface of the service is pleasant. Evaluators are neutral regarding finding the service very helpful to their work, that the service has all the functions they expected, that they will use this service in their work and that they will recommend this service to their peers/community. While comments related to the whether the service has all the functions they expected were in general positive such as *"Does what is expected, good extraction quality."* (DE3) and *"All significant functions are available"* (GR9), they also indicated some shortcomings summarized by comments such as *"It could have more functions, in order to help me find whatever I need"* (GR7) and uncertainty as exemplified by comments such as *"Not aware of list of functionalities"* (ES12) and *"I don't know what function to expect"* (GR1).

Regarding accessibility and acceptability, the evaluators agree that they are satisfied with the access given to this Dicode service, with the instruction given by this Dicode service, with the time spent to learn to use the provided solution and with the ease of use of the solution to complete their tasks. Characteristic comments were *"I learned to use the provided solution very quickly."* (GR4), *"It is easy to understand the use of this solution."* (ES15) and *"Easy and provides good results."* (UK1).

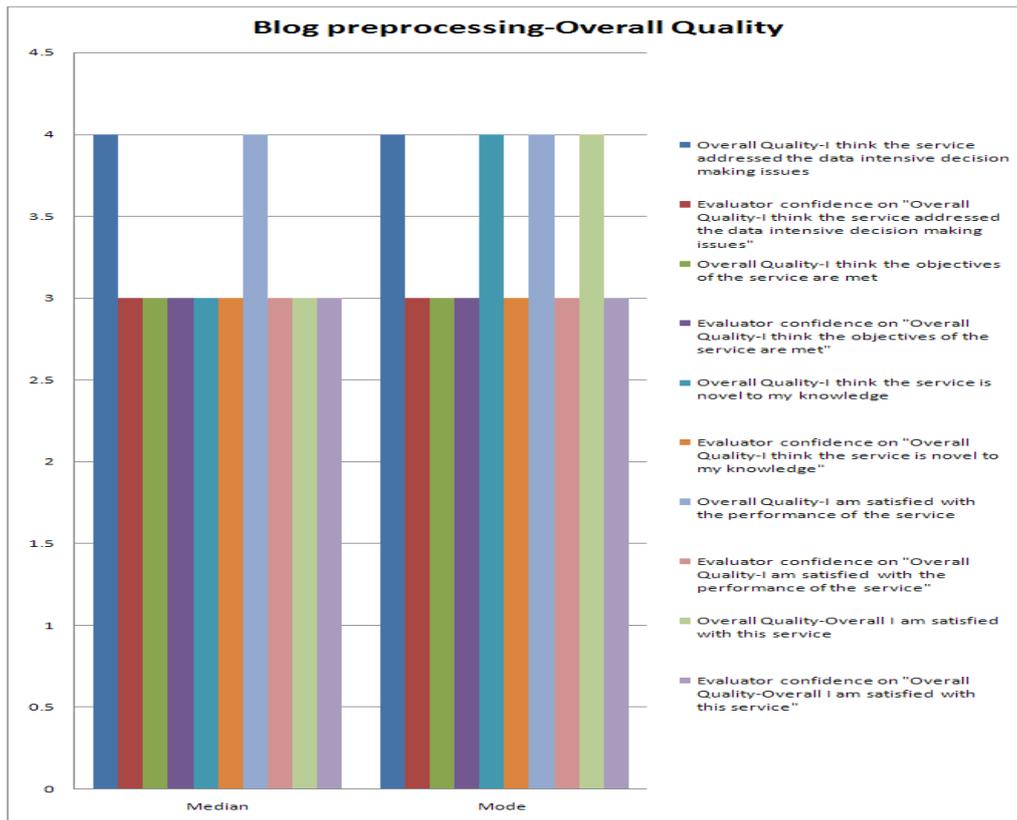


Figure 3.1 Blog pre-processing services' quality.

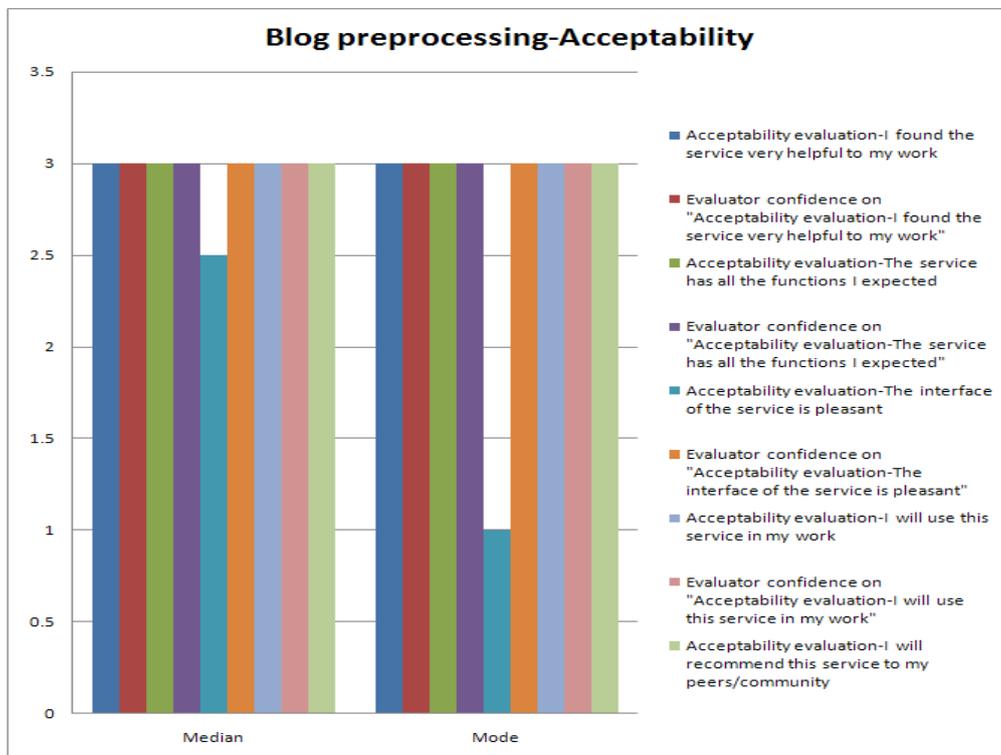


Figure 3.2 Blog pre-processing services' acceptability.

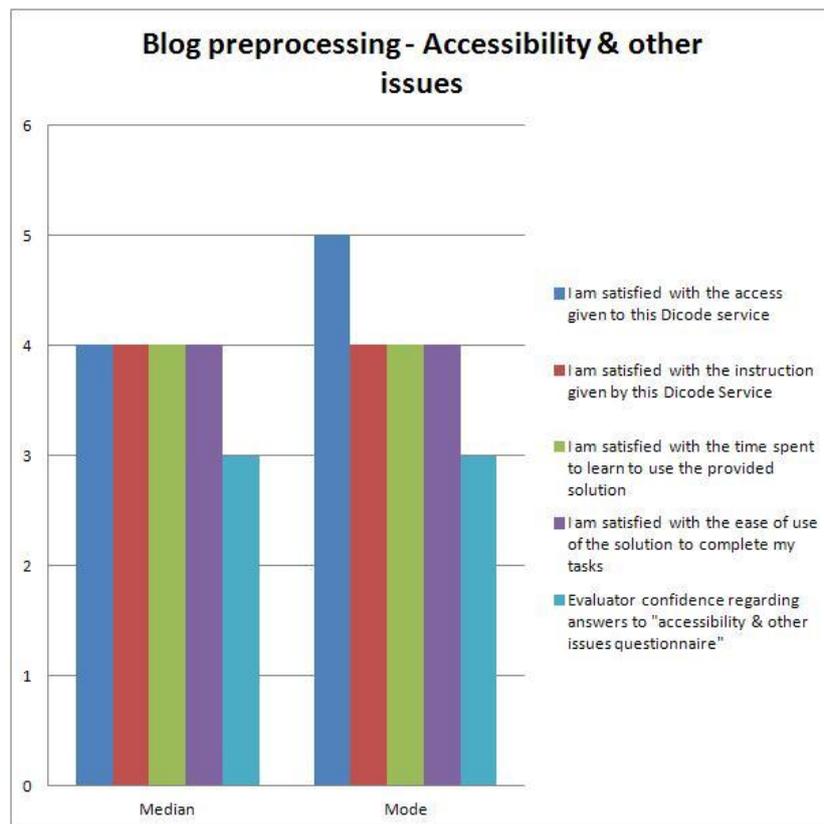


Figure 3.3 Workbench services' accessibility and other issues.

In Table 3.2, the mean values of the responses relative to usability are given (the scale in this case is 0-10). Particularly, the following usability measures are assessed: visibility, tolerance, physical mapping, restrictions, conceptual models, feedback, error prevention, flexibility, ease of recognition, flexibility of the use efficiency, provision of clear error messages, aesthetic of the minimalist design, help and documentation facilities, visibility, user control capabilities, consistency and presentation standards, fault prevention, ease of detection. Overall, the mean values are above 4.5 across categories. Especially low values are reported for the flexibility of the use efficiency: shortcuts provision/capabilities for the user to affect the configuration of the system, the fault prevention, the user control capabilities and the freedom of action, and the help/documentation facilities. The highest scores were reported for the tolerance, physical mapping, conceptual models and the consistency and presentation standards.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	49	0	10	5.22	2.946
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	49	0	10	6.51	2.526
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	49	0	10	6.31	2.830
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	49	0	10	5.39	2.797
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	49	0	10	6.08	2.507
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	49	0	10	5.43	2.880
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	49	0	10	5.31	2.647
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	49	0	10	5.18	2.539
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	49	0	10	5.92	2.581
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	49	0	10	4.80	2.806
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	49	0	10	5.45	2.475
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	49	0	10	6.00	2.558
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	49	0	10	5.12	2.659
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	49	0	9	4.86	2.677
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	49	0	10	5.59	2.605
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	49	0	10	5.10	2.939
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	49	0	10	6.10	2.823
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	49	0	9	5.08	2.597
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	49	0	10	5.96	2.776

Table 3.2. Blog pre-processing – Usability principles Descriptive Statistics

3.2.2.2 Topics service

OMUWD evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.4, 3.5 and 3.6 summarize the evaluators' responses relative to the quality, acceptability, accessibility & acceptability of the Topics service (the median values of the evaluators' responses are also shown, given in a 1-5 scale). Regarding overall quality, the evaluators agree that the service addressed the data intensive decision making issues, that the objectives of the service are met, that they are satisfied with the performance of the service, and that overall they are satisfied with this service. More specifically, characteristic comments included: *"I agree that this service addressed the data-intensive decision making issues"* (GR32), *"The service provides reasonable results and overcomes the objectives."* (UK01) and *"In my trial it was fast."* (DE01). The evaluators are neutral regarding thinking that the service is novel to their knowledge. This was reflected in comments such as *"I wouldn't consider it novel to my knowledge."* (GR15), *"There are more approaches I think."* (ES09) and *"Very common but useful for data analysis and creation of recommendations"* (GR09).

As far as the acceptability of the service is concerned, the evaluators are neutral in finding the service very helpful to their work, that the service has all the functions they expected, that the interface of the service is pleasant, that they will use this service in their work and finally that they will recommend this service to their peers/community. Comments related to these questions included suggestions to improve various aspects of the service as exemplified by comments such as *"It could have much more metadata analysis. Though, it seems useful for its task."* (GR36), *"It would be nice to have one word or sentence for a topic."* (DE02) and *"There is the right function, but I would like some more"* (GR07).

Regarding accessibility and acceptability, the evaluators agree that they are satisfied with the access given to this Dicode service, that they are satisfied with the instruction given by this Dicode service. They are satisfied with the time spend to learn to use the provided solution and with the ease of use of the solution to complete their tasks. Comments such as *"I was given enough access to complete the test."* (GR06), *"Very easy API - quick to understand"* (DE01) and *"The instructions were very good and helped me understand the service in a few minutes"* (GR12) summarize respectively the evaluator's view on the above issues.

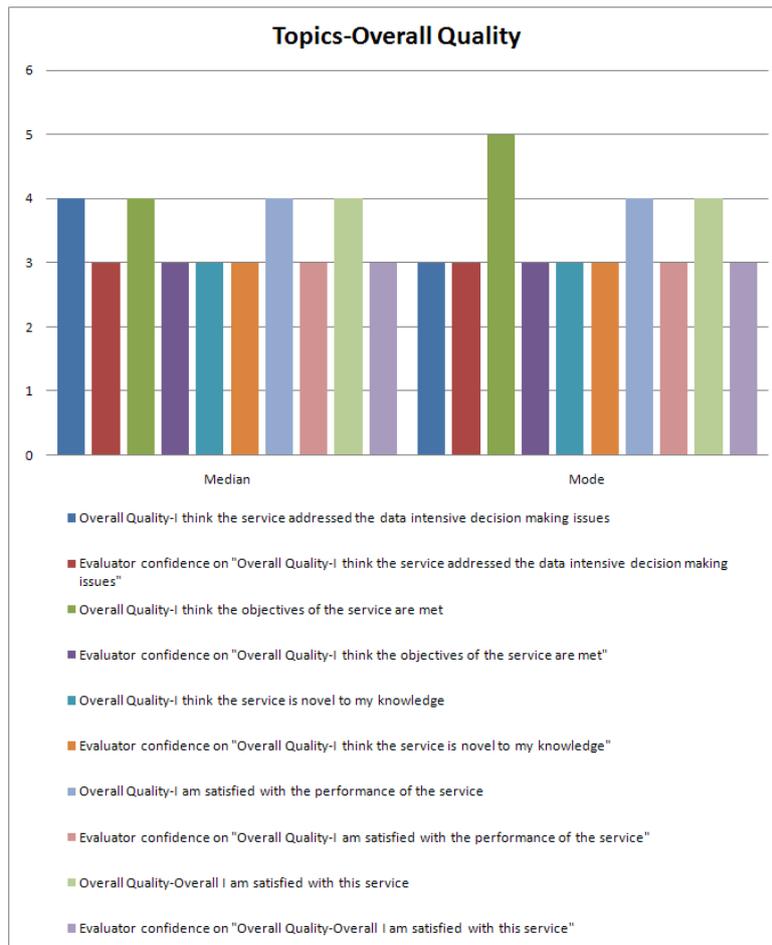


Figure 3.4 Topics service quality.

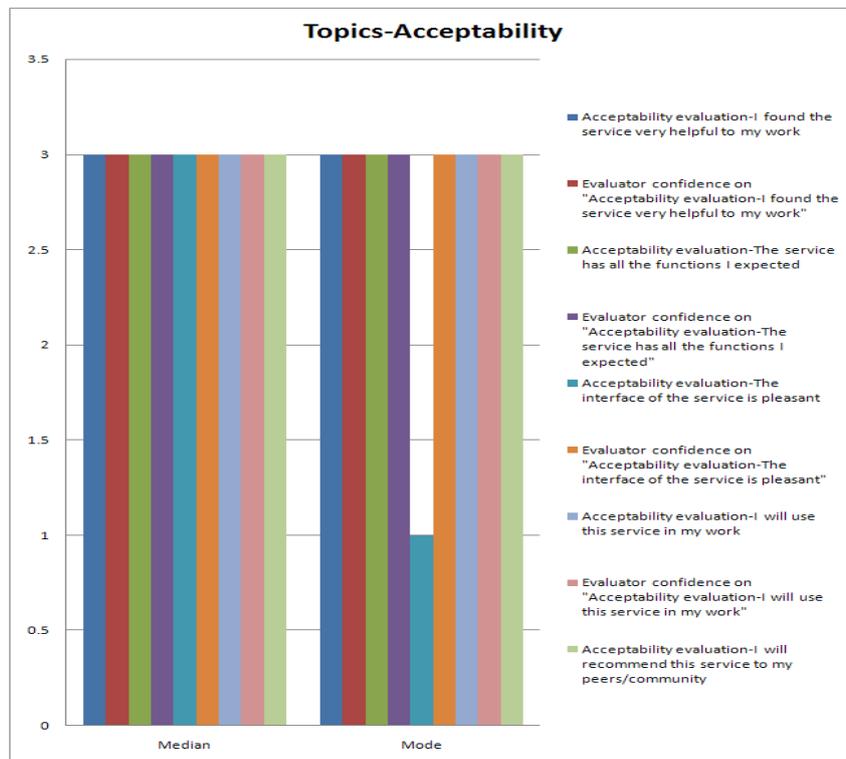


Figure 3.5 Topics service acceptability.

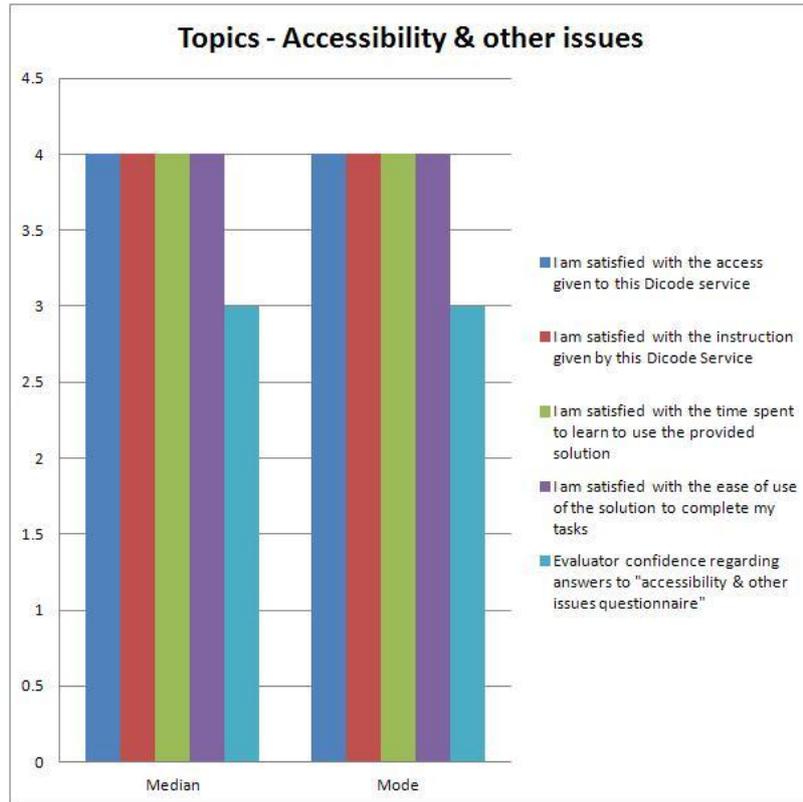


Figure 3.6 Topics service accessibility and other issues.

In Table 3.3, the mean values of the responses relative to usability are given in a 0-10 scale. Overall, the mean values are above 5. Low values were reported for fault prevention and the flexibility of the use efficiency. The highest values reported refer to the tolerance, the physical mapping and the conceptual models.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	44	0	10	5.75	2.829
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	44	3	10	7.18	2.015
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	44	0	10	6.77	2.166
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	44	0	10	5.84	2.178
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	44	1	10	6.41	2.094
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	44	0	10	5.64	2.651
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	44	0	10	5.59	2.546
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	44	0	10	5.82	2.335
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	44	0	10	6.05	2.524
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	44	0	9	5.41	2.499
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	44	0	10	5.84	2.702
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	44	0	10	5.98	2.426
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	44	0	10	5.61	2.626
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	44	0	10	5.57	2.636
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	44	1	10	6.32	2.228
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	44	0	10	5.52	2.881
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	44	0	10	6.02	2.783
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	44	0	10	5.39	2.264
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	44	0	10	6.27	2.395

Table 3.3. Topics service – Usability principles Descriptive Statistics

3.2.2.3 Keytrends service

OMUWD evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.7, 3.8 and 3.9 summarize the evaluators' responses relative to the quality, acceptability, accessibility & acceptability of the Keytrends service (the median values of the evaluators' responses are shown in a 1-5 scale). Regarding overall quality, the evaluators agree that the service addressed the data intensive decision making issues, that the objectives of the service are met, that the service is novel to their knowledge, that they are satisfied with the performance of the service and that overall they are satisfied with the service. Respective comments included *"Yes this service addressed the data-intensive decision making issues."* (GR14), *"KeyTrends always mean the most frequent. This is returned in this service."* (DE02), *"It is quite novel"* (ES03), *"The service works efficiently"* (GR02) and *"It's a simple to use and very intuitive service"* (ES11).

As far as the acceptability of the service is concerned, the evaluators are neutral that they found the service very helpful to their work, that the service has all the functions they expected, that the interface of the service is pleasant, and that they will use this service in their work. Typical comments from evaluators were *"I understand this service is very helpful, but it's not helpful to my work"* (GR32), *"It provides general statistics, these statistics are useful to me but I prefer personalized information."* (GR36) and *"The interface is pleasant, although a GUI might be easier for inexperienced users"* (GR31). Evaluators are between neutral and agree that they will recommend this service to their peers/community. Related to this questions, the relevant comments showed a great variety ranging from *"I will recommend this service to my peers/community"* (GR20), to *"Depends on too many factors"* (DE03) to *"I don't think so"* (ES20).

Regarding accessibility and acceptability, the evaluators agree that they are satisfied with the access given, the instruction given by the service, the time spend to learn to use the provided solution and for satisfaction with the ease of use of the solution to complete their tasks. Comments such as *"I am very satisfied with this Dicode service"* (GR05), *"Easy to understand instructions."* (UK01), *"Learnt very quickly"* (ES12) and *"It will be easy to incorporate into my task."* (DE02) summarize the evaluator's view related to the above issues.

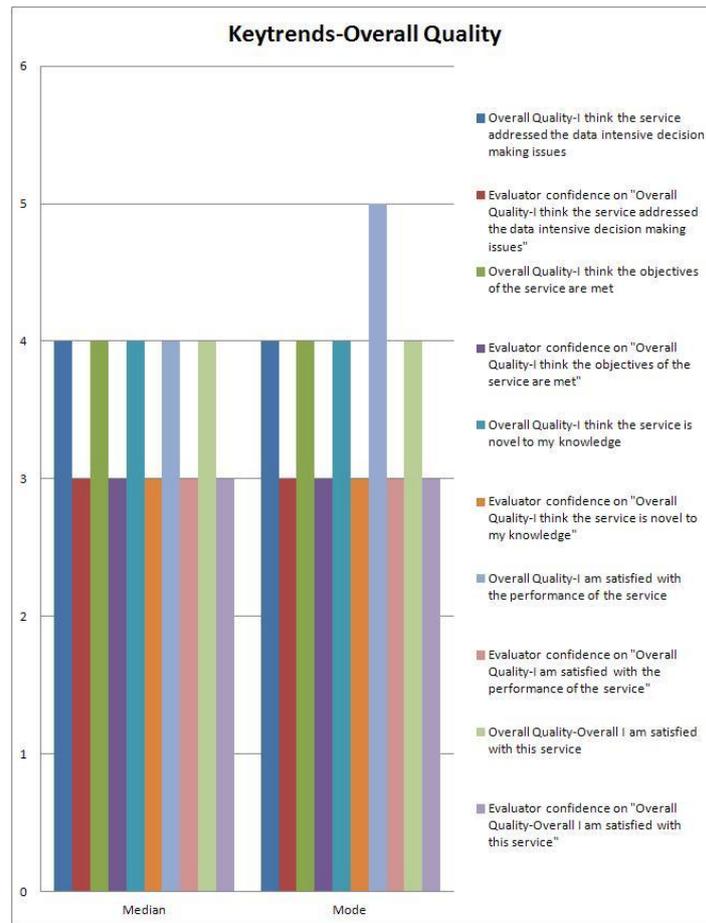


Figure 3.7 Keytrends service quality.

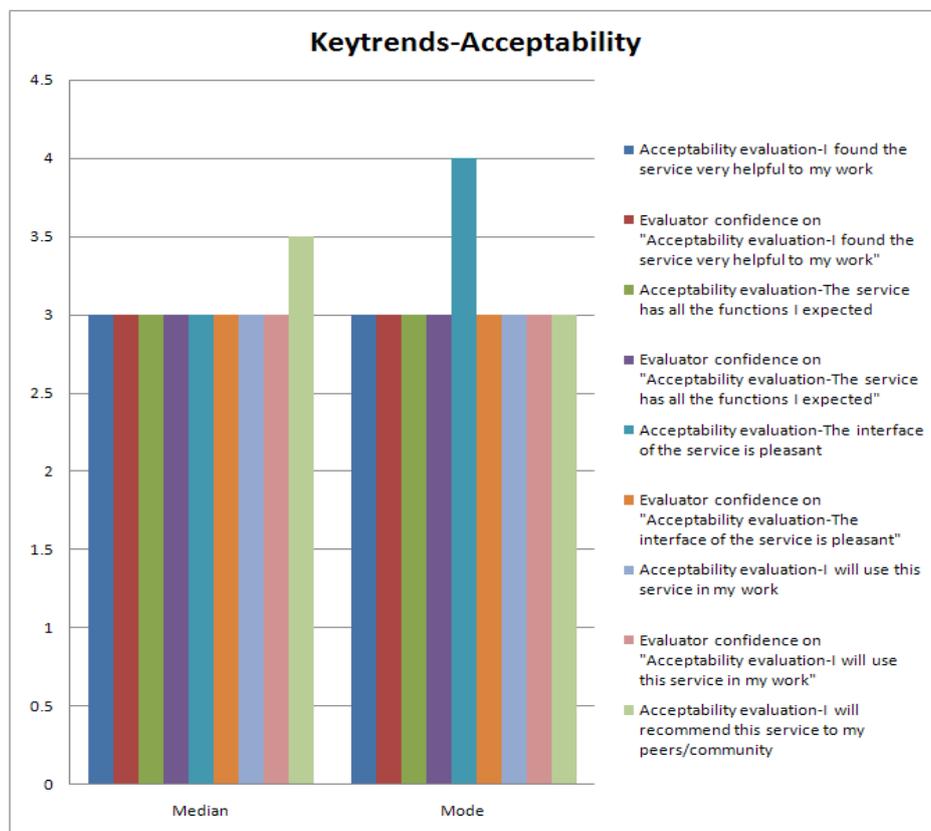


Figure 3.8 Keytrends service acceptability.

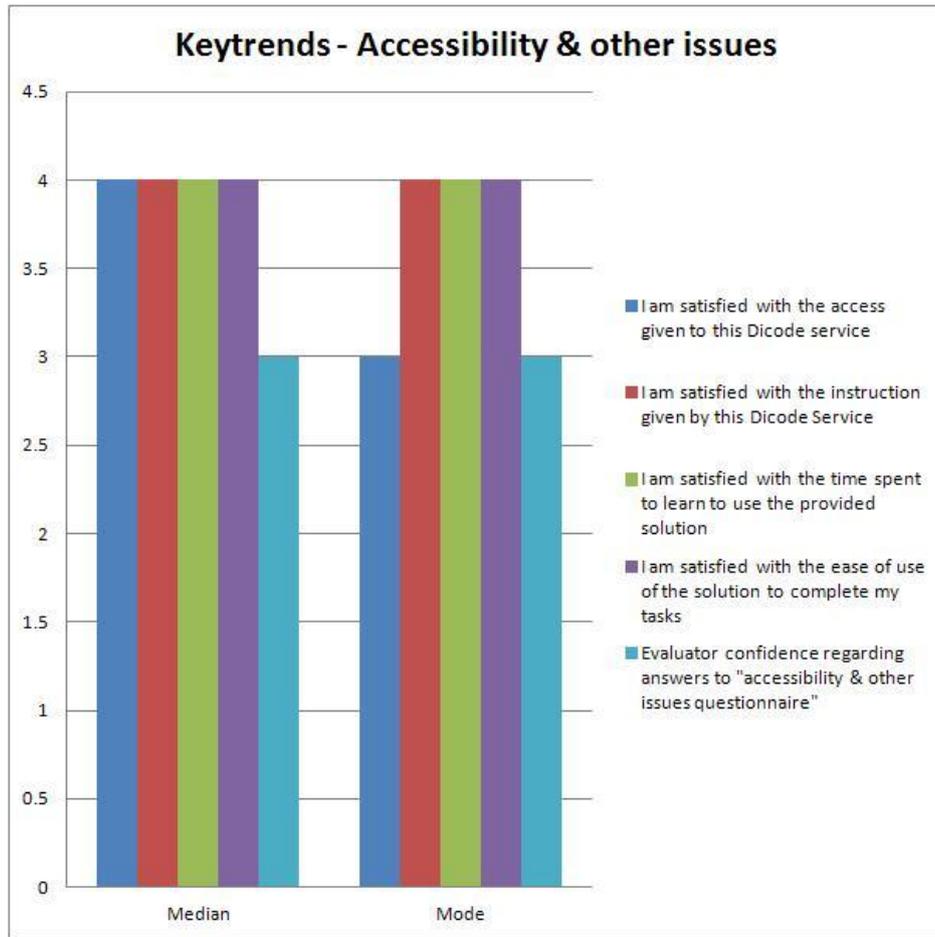


Figure 3.9 Keytrends service accessibility and other issues.

In Table 3.4, the mean values of the responses relative to usability are given; the scale in this case is 0-10. Overall, the mean values are above 4.5. Especially low scores were given for the flexibility of the use efficiency, the user control capabilities and the freedom of action and the help/documentation facilities, whereas high scores were given to the tolerance, conceptual models (proposed actions relative to perception of users for actions) and physical mapping of the service.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	49	0	10	5.41	2.943
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	49	1	10	7.33	2.240
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	49	0	10	6.65	2.876
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	49	0	10	5.82	2.430
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	49	0	10	6.76	2.204
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	49	0	10	5.24	2.673
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	49	0	10	5.53	2.639
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	49	0	10	5.16	2.577
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	49	0	10	6.06	2.742
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	49	0	9	4.80	2.723
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	49	0	10	5.22	2.867
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	49	0	10	6.06	2.617
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	49	0	10	5.02	3.024
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	49	0	10	5.12	2.595
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	49	0	9	5.94	2.427
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	49	0	10	4.90	3.022
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	49	0	10	6.10	3.043
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	49	0	10	5.69	2.785

Table 3.4. Keytrends service – Usability Principles Descriptive Statistics

3.2.2.4 Twitter pre-processing service

OMUWD evaluators used the service and completed the related Dicode Evaluation Framework Questionnaires. Figures 3.10, 3.11 and 3.12 summarize the evaluators' responses relative to the quality, acceptability, accessibility and acceptability of the Twitter pre-processing service (the median values of the evaluators' responses are shown, given in a 1-5 scale). Regarding overall quality, the evaluators are between neutral and agree that the service addressed the data intensive decision making issues and that the objectives of the service are met. Evaluator's comments indicated that *"Time of completion and use of this service are acceptable. However, my opinion is that the results are a bit complex even for an expert user."* (GR39), as well as *"The part of data searching is completed but there are no decisions taken as far as the results. Maybe the result may be given clustered."* (GR09). The evaluators are neutral when thinking about whether the service is novel and whether they are satisfied with the performance of the service. However, evaluators agreed that overall they are satisfied with the service.

Regarding acceptability, evaluators are neutral that the service is helpful to their work, that the interface of the service is pleasant, that they will use this service in their work and that they will recommend this service to their peers/community. Relevant comments indicated *"Although very easy to work with this service, it is not helpful to my work."* (GR24), *"Not my field but I can see the use of it."* (GR23), *"I haven't used the service for any work so far, but I could imagine that it might be a very good data source for social media monitoring applications"* (DE03), *"This interface is nice but I can say that it is a little bit confused for the user of this service."* (GR14), *"the huge amount of data returned makes this service usable only paired with a powerful and lightweight/fast 'data processing' application"* (GR28) and *"In this phase of development, I will not recommend this service to my peers because I believe it is a bit immature"* (GR39). The evaluators agree that the service has all the functions they expected as pointed out by comments such as *"The service has all the functions I expected."* (GR25) and *"All functions seem appropriate and useful. It might also be useful to provide a parameter to return an answer with a minimum and maximum number of nouns."* (UK1).

For accessibility and acceptability, the evaluators liked the instructions of the service and the access that was provided. They are satisfied with the time spend to learn to use the provided solution and are satisfied with the ease of use of the solution to complete their tasks. Characteristic comments include *"I didn't have any problems with accessing the Service."* (GR02), *"The instructions were useful and easy to understand."* (DE02), *"It was very easy for me to use the service. I read the manual just once and I was able to use it without any problems."* (GR12) and *"The service was straight forward to use"* (GR35).

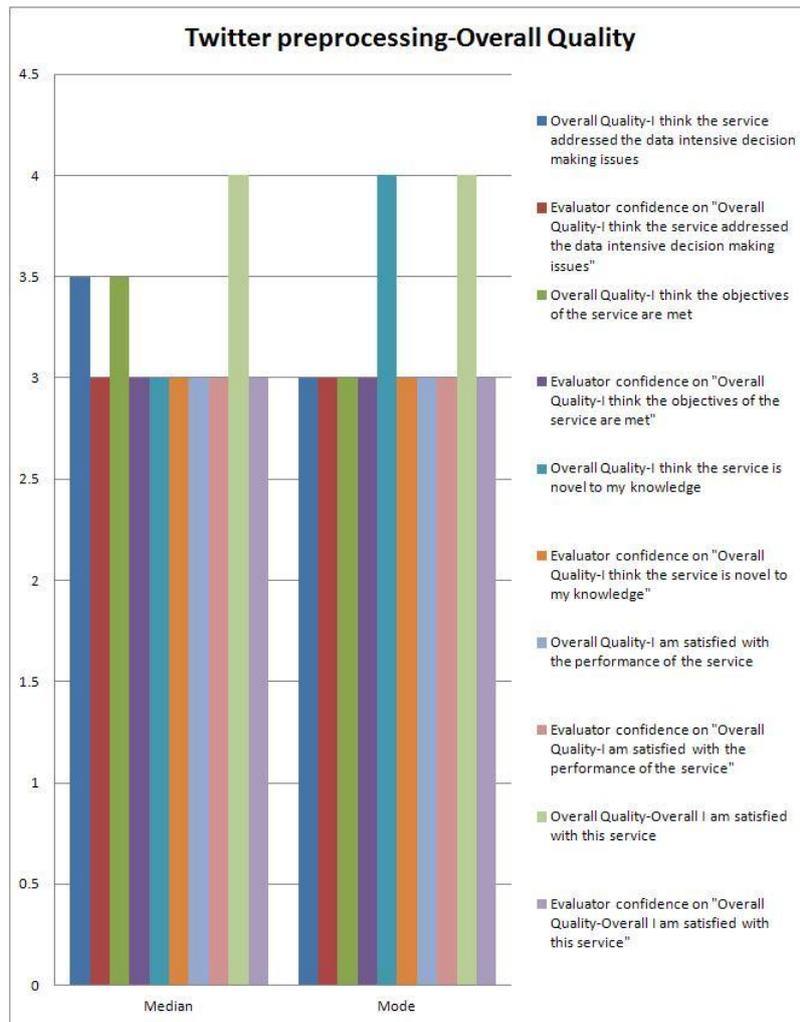


Figure 3.10 Twitter pre-processing service quality.

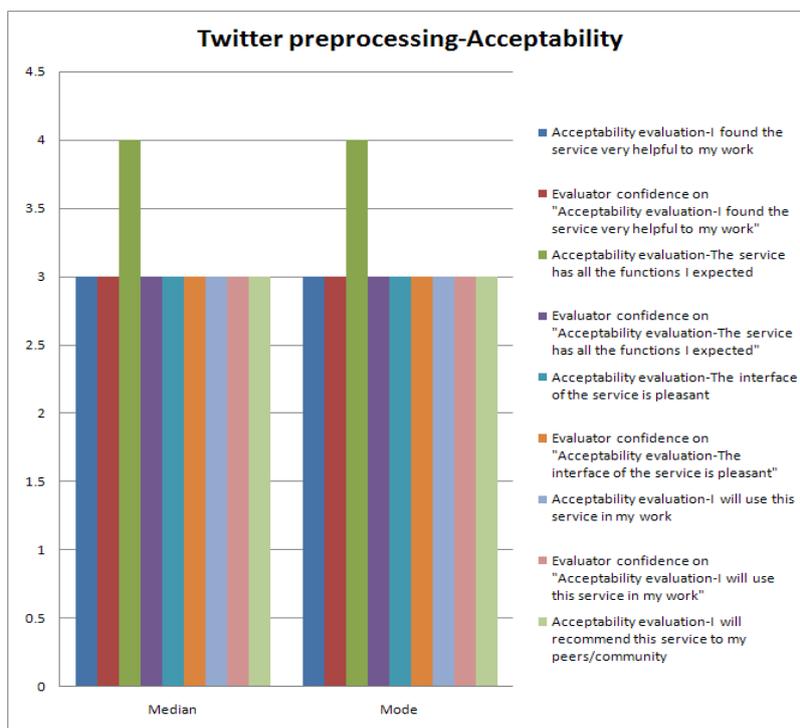


Figure 3.11 Twitter pre-processing service acceptability.

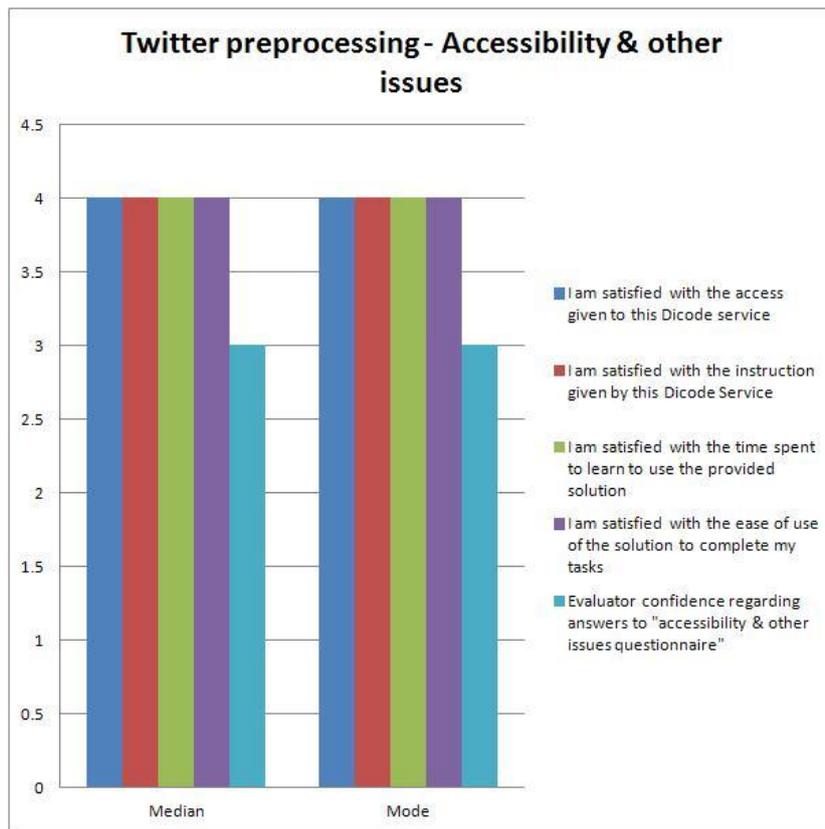


Figure 3.12 Twitter preprocessing accessibility and other issues.

In Table 3.5, the mean values of the responses relative to usability are given in 0-10 scale, where overall values are reported to be above 4.5. Particularly low values were given for the user control capabilities and the freedom of action, the provision of clear error messages and the fault prevention of the service. High values were given for the tolerance, the physical mapping, the ease of identification of the required actions and the conceptual models of the service.

Questions	N	Minimum	Maximum	Mean	Std. Deviation
Usability principles (Norman, Nielsen)-Rate the visibility: optical information to facilitate the user	56	0	10	5.71	2.921
Usability principles (Norman, Nielsen)-Rate the tolerance: behavior similar to expectations	56	0	10	7.00	2.036
Usability principles (Norman, Nielsen)-Rate the physical mapping: conceptual correspondence between commands and functions	56	0	10	6.63	2.548
Usability principles (Norman, Nielsen)-Rate the restrictions: incomplete ways in the actions performed and in the design of operations in understandable manner	56	0	10	5.75	2.671
Usability principles (Norman, Nielsen)-Rate the conceptual models: the operation of the proposed actions according to the perception of user for these actions	56	0	10	6.48	2.320
Usability principles (Norman, Nielsen)-Rate the feedback: notification regarding the user's position	56	0	10	5.70	2.669
Usability principles (Norman, Nielsen)-Rate the error prevention: restrict user errors & support for their solution	56	0	10	5.43	2.756
Usability principles (Norman, Nielsen)-Rate the flexibility: variety of operation modes	56	0	10	5.71	2.599
Usability principles (Norman, Nielsen)-Rate the ease of recognition: easy identification of the required actions	56	0	10	6.48	2.335
Usability principles (Norman, Nielsen)-Rate the flexibility of the use efficiency: shortcuts provision, capabilities for the user to affect the configuration of the system	56	0	10	5.34	3.129
Usability principles (Norman, Nielsen)-Rate the provision of clear error messages: simple language in error messages and a proposal to resolve them	56	0	10	5.02	2.870
Usability principles (Norman, Nielsen)-Rate the aesthetics of the minimalist design: messages with the necessary information	56	0	10	6.09	2.560
Usability principles (Norman, Nielsen)-Rate the help facilities and the documentation facilities: help facilities related to user action	56	0	9	5.48	2.649
Usability principles (Norman, Nielsen)-Rate the visibility of the system status: user briefing regarding the work progress	56	0	10	5.25	2.609
Usability principles (Norman, Nielsen)-Rate the system matching with the real world: comprehensive and explanatory language	56	0	10	5.75	2.739
Usability principles (Norman, Nielsen)-Rate the user control capabilities and the freedom of action: understandable and direct processes as undo and redo	56	0	10	4.91	3.005
Usability principles (Norman, Nielsen)-Rate the consistency and presentation standards: maintain the same presentation of the interface	56	0	10	5.89	3.172
Usability principles (Norman, Nielsen)-Rate the fault prevention: mechanism for preventing error occurrence	56	0	10	5.13	2.803
Usability principles (Norman, Nielsen)-Rate the ease of detection: obvious & easy actions required during service use	56	0	10	6.21	2.768

Table 3.5. Twitter pre-processing - Usability principles Descriptive Statistics

3.2.3 Reliability analysis

Reliability analysis is an assessment of the degree of consistency between multiple measurements of a variable (Kastania & Kossida, 2011). The most commonly used measure of reliability is internal consistency (i.e., consistency among the variables in a summated scale). Reliability analysis was performed using the Cronbach's α reliability coefficient (Cronbach, 1951). The Cronbach's α reliability coefficient measures the consistency among individual items in a single scale, i.e. how closely related the items of a set are as a group (Norman, 1998).

For the questionnaires examined in the context of Dicode project, it indicates how well the different items (questionnaire questions expecting a quantitative answer) complement each other in their measurement of different aspects of the same variable, i.e. reliability of service. As a rule of thumb, a value of 0.7 or higher is obtained on a substantial sample (Lewis, 1995). It is worth noting that the reliability coefficient α depends on the number of users that completed the questionnaires, and on the number of questions which expect a quantitative answer (number of items).

In Table 3.6, Cronbach α values range from 0.956 to 0.747. It is evident that the Dicode Evaluation Framework questionnaires used for each service show high internal consistency as indicated by high alpha coefficients, which exceed the recommended level of 0.7.

Dicode Service	Dicode Evaluation Framework Questionnaires	Users	Reliability Statistics	
			Cronbach's Alpha	Number of Items
Blog preprocessing API	Use Case/Blog preprocessing API Validation Instrument	50	0.928	30
	Blog preprocessing API Ease of Use Assessment	49	0.950	25
	Blog preprocessing API Accessibility & Acceptability Assessment	50	0.834	5
Topics	Use Case/Topics Service Validation Instrument	45	0.957	30
	Topics Service Ease of Use Assessment	44	0.956	25
	Topics Service Accessibility & Acceptability Assessment	45	0.868	5
Keytrends API	Use Case/Keytrends API Validation Instrument	54	0.923	30
	Keytrends API Ease of Use Assessment	49	0.943	25
	Keytrends Accessibility & Acceptability Assessment	51	0.804	5
Twitter preprocessing API	Use Case/Twitter Preprocessing Validation Instrument	54	0.916	30
	Twitter Preprocessing Ease of Use Assessment	56	0.948	25
	Twitter Preprocessing Accessibility & Acceptability Assessment	54	0.747	5

Table 3.6: Reliability Analysis Table: it shows the Cronbach α reliability statistic measures for all the CGRA Dicode services.

4 Future work directions

This deliverable reports on the first version of the evaluation of the Dicode suite of services and innovative work methodologies for OMUWD that have been designed and developed in the context of Task 6.5. Much attention was paid to the assessment of their usability and acceptability. Generally speaking, the feedback received was positive, in that it reflects the potential of the Dicode solution for the use case. Evaluators agreed that the related Dicode services are a promising solution towards dealing with unstructured Web 2.0 data but they call for improvements as far as user interfaces, documentation and functionality are concerned.

A detailed consideration of the feedback obtained from this first evaluation round will take place towards reassessing both the particular needs of the OMUWD case and the ongoing development of the related Dicode services. Accordingly, future work directions concern the fine-tuning and/or enhancement of the already implemented (and tested) services, as well as the development of additional ones, if needed.

With respect to the enhancement of already developed services, issues raised from the first evaluation process concerning the first five services mentioned in Section 2.4 are reported in deliverable D6.2.1. With respect to the last four services of Section 2.4, issues raised are:

- **Blog pre-processing service:** technical issues were raised by the reviewers relative to the service's interface and overall functionality. Many evaluators pointed out the need for a more pleasant interface and a better presentation of the service's functionalities.
- **Topics service:** Similar to the issues raised for the service above. Moreover, evaluators wanted to see a more elaborated metadata analysis, as well as a better documentation of the available functionality.
- **Keytrends service:** In addition to the issues mentioned above, evaluators asked for improvements in terms of information personalization.
- **Twitter pre-processing service:** Evaluators mainly pointed out the need to improve the service's user interface towards further facilitating sense-making with respect to the service's outcomes. Also, they recommended the interoperation of this service with other – case related – data processing applications.

As far as the feedback obtained by the interviews with the marketing professionals is concerned, we claim that it was very positive. The Dicode Workbench approach seems to cover their requirements in terms of data and services integration. The overall approach followed in Dicode with respect to collaboration and decision making is fully in line with their needs for virtual workspaces, which can be of different formality and concern different stakeholders. The alternative views already offered by the Dicode's Collaboration Support Services are in the right direction. Future work directions indicated by these professionals concern the provision of private workspaces, the inclusion of additional (or enhancement of the already provided) views that further augment sense-making (e.g. collaboration summaries), as well as the capability of integrating the Dicode services with third-party applications (e.g. Customer Relationship Management Systems).

References

- Baker, M. (2012). Quantitative data: learning to share. *Nat Meth* , 9, 39-41.
- Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika* , 16 (2), 297-334.
- Crockford, D.: JSON: The fat-free alternative to XML. In: Proc. of XML 2006, Boston, USA (December 2006), HYPERLINK "http://www.json.org/fatfree.html"
<http://www.json.org/fatfree.html>
- Dicode Deliverable D2.2. (2011). *The Dicode Approach User requirements, conceptual integrative architecture, agile methodology and functional specifications*.
- Dicode Deliverable D3.1.1. (2011). *The Dicode Data Mining Framework (initial version)*.
- Dicode Deliverable D3.2.1. (2011). *The Dicode Data Mining Services (initial version)*.
- Dicode Deliverable D4.1.1. (2011). *The Dicode Collaboration Support Services (initial version)*.
- Dicode Deliverable D4.2.1. (2011). *The Dicode Decision Making Support Services (initial version)*.
- Dicode Deliverable D5.4.1. (2011). *Integrated Dicode Services (initial version)*.
- Dicode Deliverable D6.1. (2011). *The Dicode Evaluation Framework*.
- Kastania A.N. & Kossida, S. (2011). Statistics in Telemedicine, Chapter 9 in "Advances in Telemedicine: Technologies, Enabling Factors and Scenarios" (Eds. Georgi Graschew and Theo A. Roelofs), ISBN 978-953-307-159-6, pp. 191-210, INTECH Open.
- Kitchenham, B. & Pfleeger, S.L. (2002), Principles of Survey Research Part 5: Populations and Samples, *Software Engineering Notes* vol 27 no 5, 17-20, ACM SIGSOFT
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *Hum-Comput Interact* , 7 (1), 57-78.
- Nielsen, J. (1991). *Designing Web Usability: The Practice of Simplicity*. Indianapolis: New Riders Publishing.
- Norman, D. A. (1998). *The Design of Everyday Things*. The MIT Press.

Appendix: Evaluator's Guide

We hereby include the parts of the Dicode Evaluator's Guide that concern the last four Dicode services evaluated for the OMUWD use case (see Section 2.4). Parts concerning the first five Dicode services have been included in the Appendix of D6.2.1. The guide includes general information about the service, usage scenarios which emphasize the settings under which the service is applied, and a detailed description of the actions the evaluator needs to perform during evaluation.

THE DICODE DATA MINING SERVICES

In order to use the service a REST-Client such as RESTClient (<https://github.com/chao/RESTClient>) must be installed.

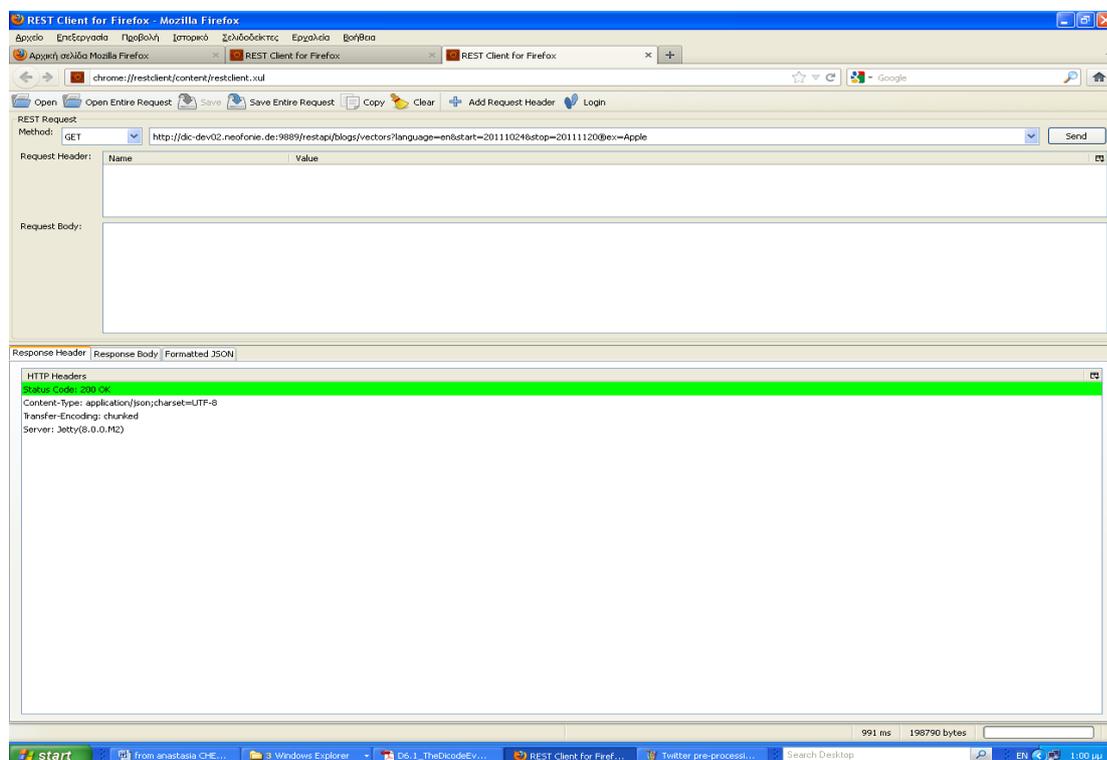
Additional instructions for each of the services below appear in:

<https://wiki.dicode-project.eu/display/DIC/Evaluation+of+Dicode+Services>

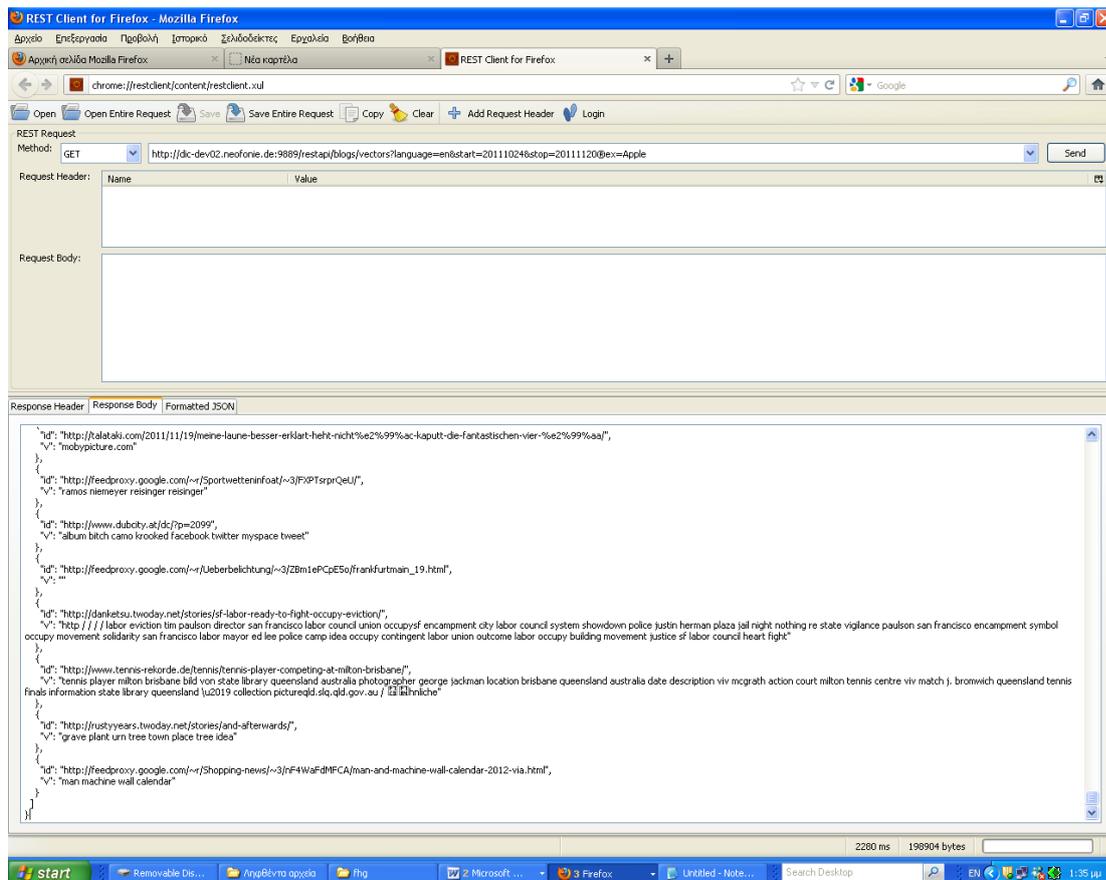
(i) Blog pre-processing API

Example Request

GET <http://dic-dev02.neofonie.de:9889/restapi/blogs/vectors?language=en&start=20111024&stop=2011120®ex=Apple>



Expected Output



Please complete the following Questionnaires:

- Dicode Evaluation - First Trial - Standalone Service - Blog pre-processing API

http://kwiksurveys.com?s=LKLJFL_93fd0847

- Dicode Evaluation - First Trial - Standalone Service - The Blog pre-processing API - Ease of Use Assessment User Questionnaire

http://kwiksurveys.com?s=LKLJFK_d999de4

- Dicode Evaluation - First Trial - Standalone Service - The Blog pre-processing API - Acceptability Assessment User Questionnaire

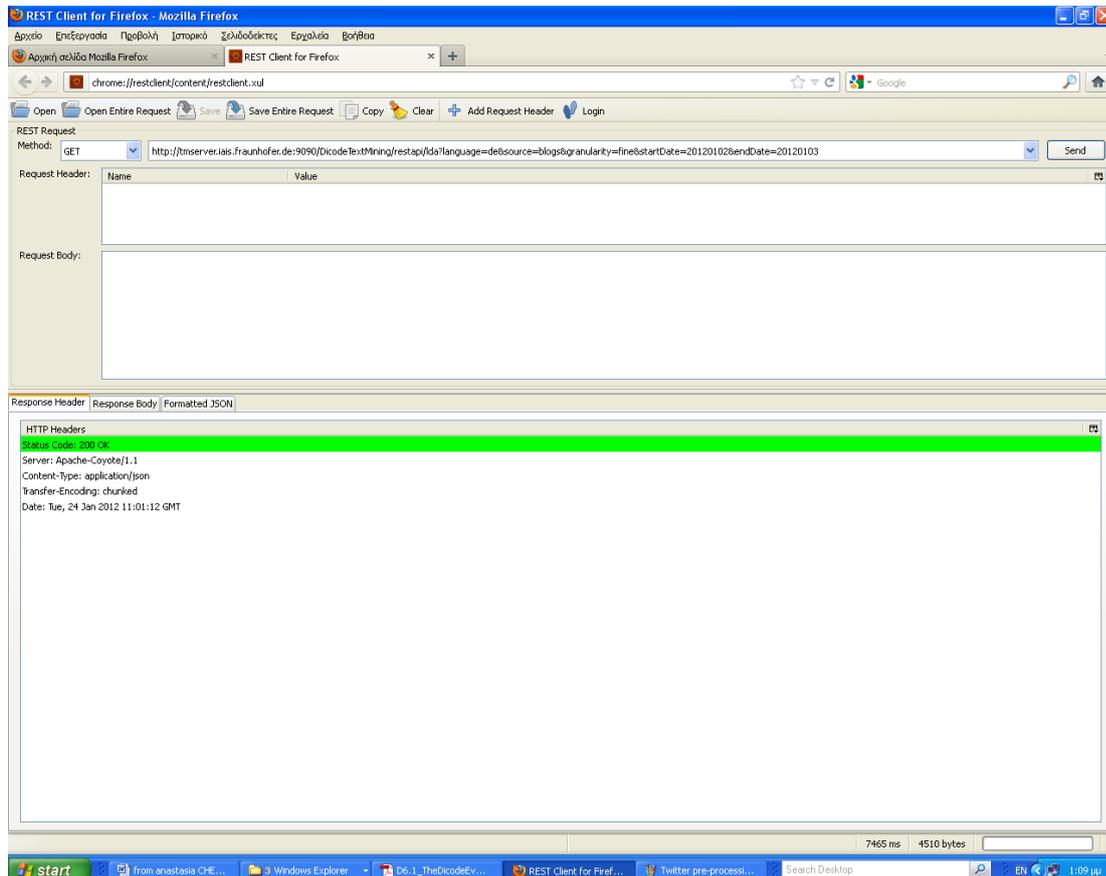
http://kwiksurveys.com?s=LKLJFJ_7a9ead72

(ii) Topics Service API

Example Request

GET

<http://tmserver.iais.fraunhofer.de:9090/DicodeTextMining/restapi/lda?language=de&source=blogs&granularity=fine&startDate=20120102&endDate=20120103>



Expected Output

The screenshot shows a REST Client for Firefox window. The URL bar contains the URL: `http://[msserver.iars.fraunhofer.de:9090]/DicodeTextMining/restapi/ids?language=de&source=blogs&granularity=fine&startDate=20120102&endDate=20120103`. The request method is GET. The response body is a JSON array of terms, each with a probability and a list of document IDs. The response is displayed in the 'Response Body' tab, formatted as JSON.

```

{"Number of topics":7,"Number of documents":735,"List of topics":[{"Number of relevant documents":328,"Document percentage":44.6927374301676,"Terms":[{"Term":"eur","Probability":0.024875621890547265}, {"Term":"bestseller","Probability":0.013266998314525209}, {"Term":"liste","Probability":0.0128524046434942}, {"Term":"informationen","Probability":0.012023217247097845}, {"Term":"amazon.de","Probability":0.00995024864756218905}, {"Term":"rangposition","Probability":0.009790248756218905}, {"Term":"produkte","Probability":0.009535655059043118}, {"Term":"angebote","Probability":0.00912106135986733}, {"Term":"bathn","Probability":0.008291873963515755}, {"Term":"kategorie","Probability":0.00787728026335967}, {"Term":"relevant documents":302,"Document percentage":41.11922141192214,"Terms":[{"Term":"jahr","Probability":0.0244919786962667}, {"Term":"euro","Probability":0.0256644191978697}, {"Term":"januar","Probability":0.008823529411764706}, {"Term":"blog","Probability":0.008021390374331552}, {"Term":"weiterlesen","Probability":0.007754010695187166}, {"Term":"jahre","Probability":0.006417112299465241}, {"Term":"video","Probability":0.006417112299465241}, {"Term":"artikel","Probability":0.005614973262032086}, {"Term":"gutschein","Probability":0.005614973262032086}, {"Term":"relevant documents":245,"Document percentage":33.33333333333333,"Terms":[{"Term":"papendaf","Probability":0.03130240357741755}, {"Term":"genetide","Probability":0.01676914773616545}, {"Term":"kameraden","Probability":0.009502515371716043}, {"Term":"feuerwehr","Probability":0.009502515371716043}, {"Term":"uhr","Probability":0.008943543879262158}, {"Term":"pfeilshow","Probability":0.00866405813303215}, {"Term":"tief","Probability":0.007546115148127445}, {"Term":"jahr","Probability":0.00670765790946618}, {"Term":"abschluss","Probability":0.006428172163219676}, {"Term":"emsa.de","Probability":0.006148689416992733}, {"Term":"relevant documents":159,"Document percentage":21.632766233716043,"Terms":[{"Term":"spiel","Probability":0.024793389429732067}, {"Term":"fame","Probability":0.0133805383510429}, {"Term":"bids","Probability":0.012970129070129069}, {"Term":"uhr","Probability":0.0125947138921694}, {"Term":"house","Probability":0.0114122699647777}, {"Term":"stadion","Probability":0.00944510034191263}, {"Term":"minuten","Probability":0.00944510034191263}, {"Term":"stimmung","Probability":0.00905155450609996}, {"Term":"testock","Probability":0.00905155450609996}, {"Term":"platz","Probability":0.00905155450609996}, {"Term":"relevant documents":127,"Document percentage":17.23956633005233,"Terms":[{"Term":"zeit","Probability":0.010588235294117647}, {"Term":"tag","Probability":0.010588235294117647}, {"Term":"quade","Probability":0.007352941176470588}, {"Term":"jahr","Probability":0.007352941176470588}, {"Term":"anna","Probability":0.007352941176470588}, {"Term":"gold","Probability":0.007352941176470588}, {"Term":"haus","Probability":0.006764705882352941}, {"Term":"geschenk","Probability":0.006764705882352941}, {"Term":"lante","Probability":0.006176470588235294}, {"Term":"tage","Probability":0.0058823529411764705}, {"Term":"relevant documents":273,"Document percentage":37.19512195121951,"Terms":[{"Term":"mensch","Probability":0.009590061182306739}, {"Term":"jesus","Probability":0.00702769795618024}, {"Term":"feuerwehr","Probability":0.005374121537825546}, {"Term":"fremder","Probability":0.005374121537825546}, {"Term":"lud","Probability":0.004907273737429}, {"Term":"management","Probability":0.00454733608292095}, {"Term":"amazon","Probability":0.00454733608292095}, {"Term":"dies","Probability":0.00413393964448119}, {"Term":"bids","Probability":0.00413393964448119}, {"Term":"dvd","Probability":0.00413393964448119}, {"Term":"relevant documents":236,"Document percentage":32.17391304347826,"Terms":[{"Term":"gold","Probability":0.02156796002104155}, {"Term":"ankauf","Probability":0.015781167806417674}, {"Term":"schmuck","Probability":0.01262493424513414}, {"Term":"uhren","Probability":0.01209889531825355}, {"Term":"reisen","Probability":0.0094670063350305}, {"Term":"the","Probability":0.0094670063350305}, {"Term":"amband","Probability":0.00894261756970016}, {"Term":"http","Probability":0.00894261756970016}, {"Term":"paket","Probability":0.00841662839309426}, {"Term":"dynamen","Probability":0.007364497626246}], "Parameters":{"Granularity of topic analysis":"fine","Language":"de","Document source":"blogs","Start date":"20120102","End date":"20120103"}]}

```

Please complete the following Questionnaires

- Dicode Evaluation - First Trial - Topics Service API

http://kwiksurveys.com?s=LKLELN_8c40c6dc

- Dicode Evaluation - First Trial - Standalone Service - Topics Service - Ease of Use Assessment User Questionnaire

http://kwiksurveys.com?s=LKLELM_15499766

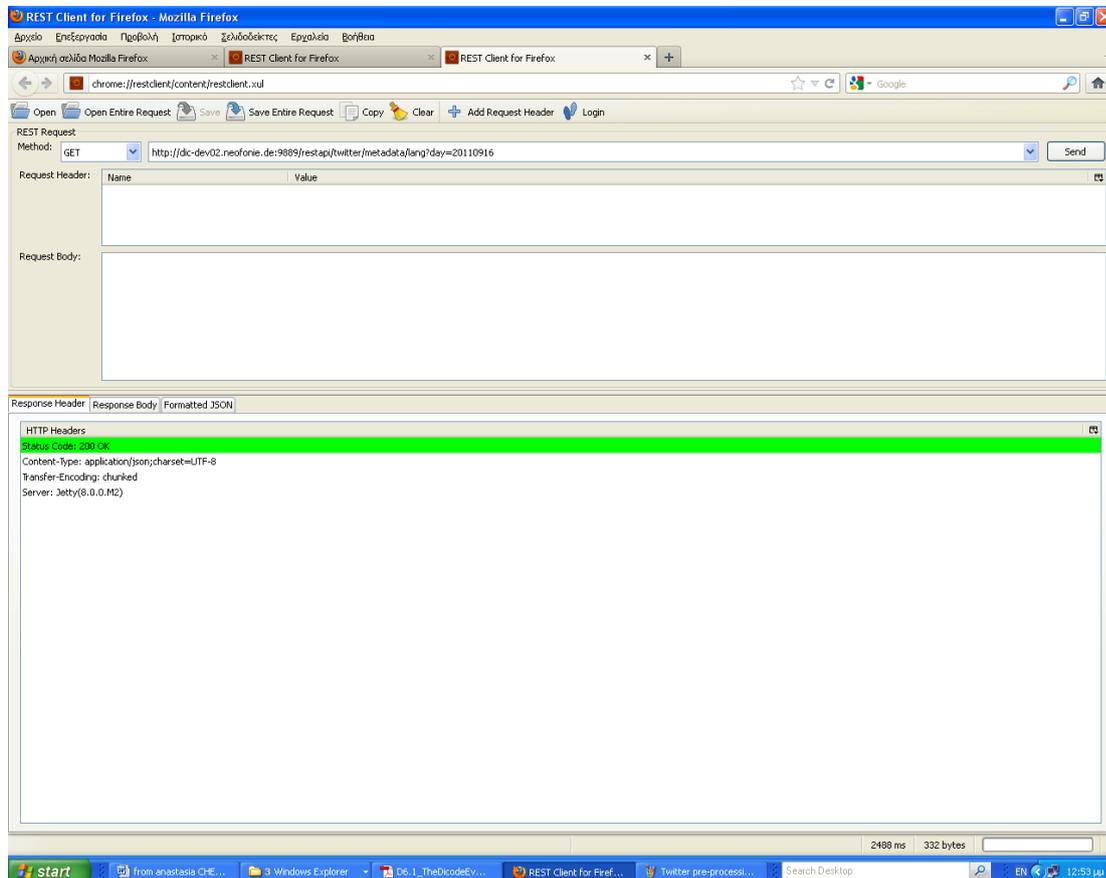
- Dicode Evaluation - First Trial - Standalone Service - Topics Service API - Acceptability Assessment User Questionnaire

http://kwiksurveys.com?s=LKLELL_624ea7f0

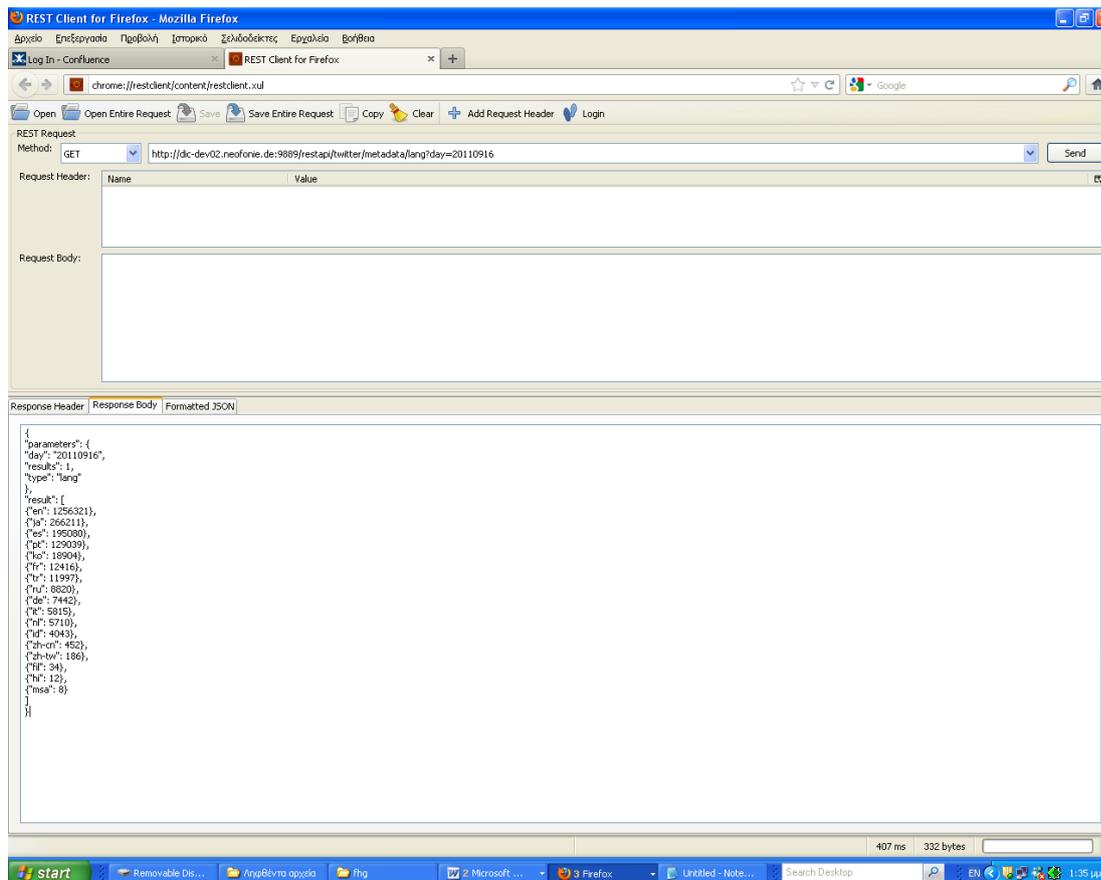
(iii) Keytrends API

Example Request

GET <http://dic-dev02.neofonie.de:9889/restapi/twitter/metadata/lang?day=20110916>



Expected Output



Please complete the following Questionnaires

- Dicode Evaluation - First Trial - Standalone Service - The Keytrends API

http://kwiksurveys.com?s=LKLJHF_edabccd7

- Dicode Evaluation - First Trial - Standalone Service - The KeyTrends API - Ease of Use Assessment User Questionnaire

http://kwiksurveys.com?s=LKLJIO_8d6c4532

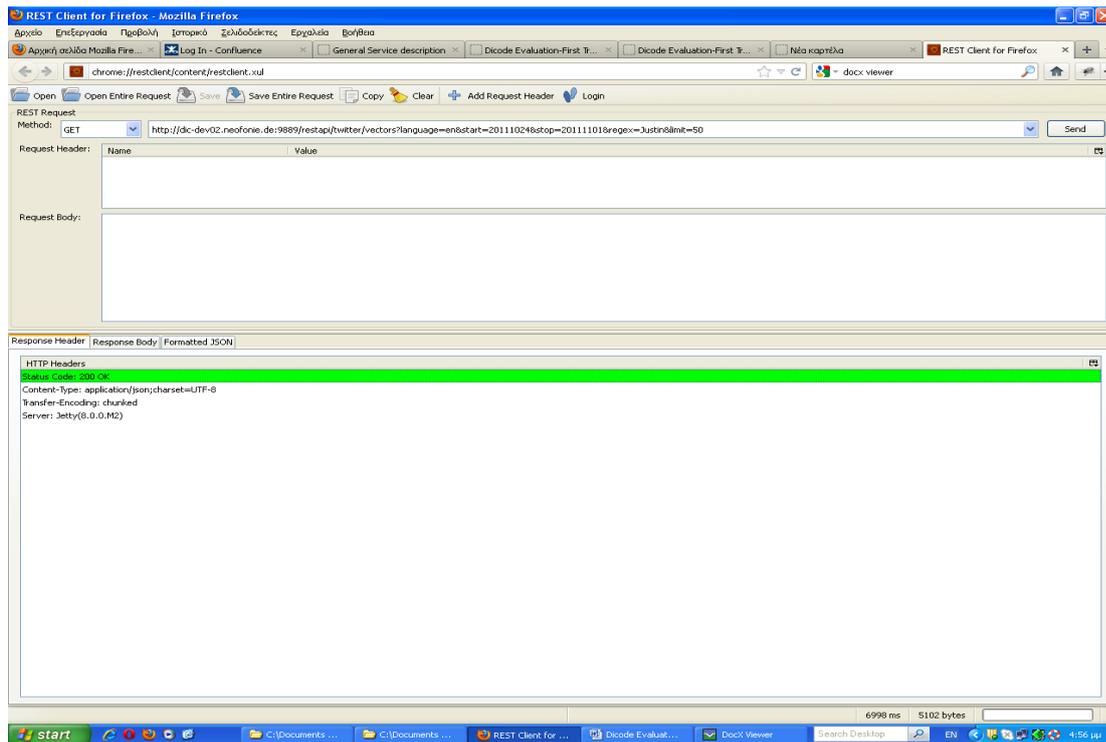
- Dicode Evaluation - First Trial - Standalone Service - The Keytrends API - Acceptability Assessment User Questionnaire

http://kwiksurveys.com?s=LKLJIL_14651488

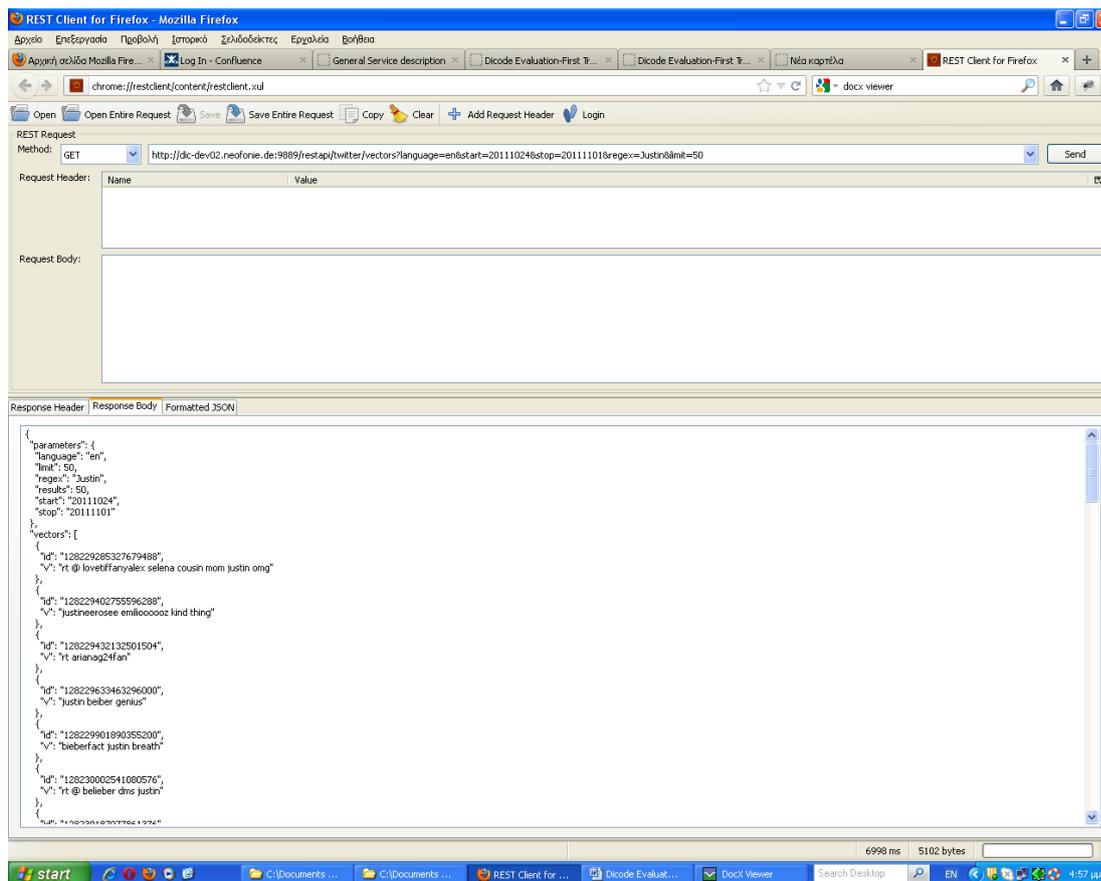
(iv) Twitter pre-processing API

Example Request

GET <http://dic-dev02.neofonie.de:9889/restapi/twitter/vectors?language=en&start=20111024&stop=20111101®ex=Justin&limit=50>



Expected Output



Please complete the following Questionnaires:

- Dicode Evaluation - First Trial - Standalone Service - Twitter pre-processing API
http://kwiksurveys.com?s=LKLKHM_7bbb7f68
- Dicode Evaluation - First Trial - Standalone Service - Twitter Preprocessing API - Ease of Use Assessment User Questionnaire
http://kwiksurveys.com?s=LKLJIG_a89a9ec3
- Dicode Evaluation - First Trial - Standalone Service - Twitter pre-processing API - Acceptability Assessment User Questionnaire
http://kwiksurveys.com?s=LKLJKM_5154469c