

MULTISENSOR

Mining and Understanding of multilingual contenT for Intelligent Sentiment
Enriched coNtext and Social Oriented inteRpretation

FP7-610411

D8.5

Final System Evaluation Report

Dissemination level:	Public
Contractual date of delivery:	Month 36, 31/10/2016
Actual date of delivery:	Month 37, 10/11/2016
Workpackage:	WP8 Use Cases and Evaluation
Task:	T8.3 Research and system quality assurance T8.4 Metrics development for theoretical solutions T8.5 User validation and prototype evaluation
Type:	Report
Approval Status:	Final Draft
Version:	1.0
Number of pages:	122
Filename:	D8.5_FinalSystemEvaluationReport_2016-11-10_v1.0.pdf

Abstract

This document contains the evaluation report with regard to the final MULTISENSOR System. It details the evaluation methodology and tools, the metrics for technical solutions, the evaluation process itself and the evaluation results. It also includes an analysis of the evaluation results and a final assessment of the MULTISENSOR system at the end of the project.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.



Co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	22.07.2016	Document initiation	N. Heise (DW)
0.2	26.09.2016	Second version	M. Puigbo (PIMEC)
0.3	24.10.2016	First consolidated version with contributions from all the partners	M. Puigbo (PIMEC)
0.4	29.11.2016	Consolidated version including input from partners	M. Puigbo (PIMEC)
0.5	3.11.2016	Internal review	G. Casamayor (UPF)
1.0	4.11.2016	Final Version for submission	M. Puigbo (PIMEC)

Author list

Organisation	Name	Contact Information
DW	Nico Heise	nicolaus.heise@dw.com
PIMEC	Martí Puigbó	mpuigbo@pimec.org
PR	Mirja Eckhoff	mirja.eckhoff@pressrelations.de
PR	Leszek Blacha	leszek.blacha@pressrelations.de
CERTH	Stefanos Vrochidis	stefanos@iti.gr
CERTH	Ioannis Kompatsiaris	ikom@iti.gr
CERTH	Dimitrios Liparas	dliparas@iti.gr
EVERIS	Andriy Bilous	andriy.bilous@everis.com
EURECAT	Ioannis Arapakis	ioannis.arapakis@eurecat.org
ONTO	Boyan Simeonov	boyan.simeonov@ontotext.com
LT	Reinhard Busch	r.busch@linguatec.de

EXECUTIVE SUMMARY

This deliverable presents the evaluation results of the final evaluation of the MULTISENSOR prototype that has been completed in September 2016.

It describes the user-centred evaluation methodology that is tailored for each use case scenario and that utilises one-to-one interviews as well as focus group interviews based on a standard questionnaire. In this final round, the evaluation followed the principles of summative testing with regard to the finished system as a whole. This third and final evaluation round also included a remote online evaluation by external partners and especially members of the MULTISENSOR user group.

The evaluation itself has been conducted by the user partners Deutsche Welle, pressrelations and PIMEC. The main features evaluated were the overall system usability and how the MULTISENSOR system helps fulfilling the different tasks that are typical for the three different use cases.

Overall, user feedback has been very positive for all three use cases. Generally, all the requirements have been implemented into the different platforms. Particularly, specific features such as summarisation, translation and decision support showed promising results and have been mentioned by the users as potentially exploitable modules. Regarding the system's interface, there was a general improvement and the users found it easy to use and navigate through.

This deliverable presents the good results of the final (summative) evaluation round. The system as such was judged as useful for the different professional tasks and the consortium received useful feedback on exploitation possibilities.

The evaluation of the Final MULTISENSOR System has followed the principles of *summative testing*. Nevertheless, in order to be able to compare the results of this summative evaluation with the results of previous evaluation rounds, the evaluation of the Final System has significant overlaps with the evaluation of the First and the Second Prototype. Consequently, in several cases this deliverable D8.5 refers to D8.3 (First Prototype Evaluation Report) and D8.4 (Second Prototype Evaluation Report) or - for better understanding - even replicates some of the statements and wording from D8.3 and/or D8.4.

Abbreviations and Acronyms

EURECAT	Eurecat
BSH	Bosch Siemens Hausgeräte GmbH
CAP	Content Alignment Pipeline
CEP	Content Extraction Pipeline
Dx.y	Deliverable x.y
DoW	Description of Work
DW	Deutsche Welle
EUMSSI	Event Understanding through Multimodal Social Stream Interpretation
FP7	7th Framework Programme
GUI	Graphical User Interface
ISO	International Organization for Standardization
MS	Milestone
NE	Named Entity
PDF	Portable Document Format
PIMEC	Petita i Mitjana Empresa de Catalunya
PPT	Microsoft PowerPoint
PUC	Pilot Use Case
PR	pressrelations GmbH
R&D	Research and Development
SME	Small or Medium Enterprise
SUG	Super User Group
Tx.y	Task x.y
URL	Uniform Resource Locator
UX	User Experience
USP	Unique Selling Proposition
WP	Work Package
WT	Workplan Table
WWF	World Wildlife Fund
XLS	Microsoft Excel

Table of Contents

1. SYSTEM DEVELOPMENT, USER EVALUATION PLAN AND EVALUATION METRICS.....	8
1.1. Background.....	8
1.1.1. Three Use Cases	8
1.1.2. User Evaluation	8
1.1.3. Formative and Summative Testing find	8
1.2. Usability Testing at MULTISENSOR	9
1.2.1. General Principles	9
1.2.2. Summative Usability Testing (Final System)	10
1.3. Evaluation metrics for theoretical solutions.....	13
2. EVALUATION RESULTS.....	15
2.1. Pilot Use Case 1: Journalism	15
2.1.1. Prototype Description and Features	15
2.1.2. Evaluation Set-Up.....	18
2.1.3. Task-related evaluation.....	18
2.1.4. Usability testing.....	21
2.1.5. General Comments	23
2.1.6. Evaluation Analysis and Final Assessment	23
2.2. Pilot Use Case 2: Commercial Media Monitoring.....	24
2.2.1. Prototype Description and Features	24
2.2.2. Evaluation Set-Up.....	29
2.2.3. Task-Related Evaluation	30
2.2.4. Usability Testing	36
2.2.5. General Comments	39
2.2.6. Evaluation Analysis and Final Assessment	39
2.3. Pilot Use Case 3: SME Internationalisation.....	41
2.3.1. Prototype Description and Features	41
2.3.2. Evaluation Set-Up.....	44
2.3.3. Task-related evaluation.....	44
2.3.4. Usability testing.....	47

2.3.5. General Comments	48
2.3.6. Evaluation Analysis and Final Assessment	48
3. FINAL ASSESSMENT OF THE MULTISENSOR SYSTEM AND THE PROJECT RESULTS.....	50
3.1. Status Assessment of the Final MULTISENSOR System	50
3.2. Assessment of the Project Results	50
4. CONCLUSION	52
5. REFERENCES.....	53
APPENDICES	54

1. SYSTEM DEVELOPMENT, USER EVALUATION PLAN AND EVALUATION METRICS

1.1. Background

1.1.1. Three Use Cases

The project has established *three* pilot use cases:

Journalism(PUC1)

Commercial media monitoring (PUC2)

SME internationalisation (PUC3)

Despite several overlaps between the use cases, requirements, interfaces and target groups show considerable differences. Consequently, evaluation structure and tasks have been tailored for each specific use case scenario. Each user partner was responsible for carrying out the evaluation for their use case. The three different pilot use cases are defined in deliverable D8.2.

1.1.2. User Evaluation

Nevertheless, user evaluation in each of the three use cases is more or less following the same principles and methodology. The general approach is a user-centred evaluation that emphasises on the role of the user rather than the system and considers the needs and limitations of the end-users. The focus lies in testing the system and specific modules in a near-real-life scenario, by giving test persons realistic tasks in a staged, but nevertheless, realistic environment. The ultimate goal of all evaluation activities is to assess the usability of the MULTISENSOR system.

1.1.3. Formative and Summative Testing find

The evaluation of the First and the Second Prototype has followed the principles of formative testing. *Formative testing* is very relevant during the development phase and focuses on identifying and fixing problems. The goal in these evaluation rounds was to provide developers with insight on how users evaluate a specific status of the prototype within the development cycle.

In contrast to this, the evaluation of the Final MULTISENSOR System was *summative*. Summative testing does not aim at supporting further development but instead seeks to assess whether the finished system as a whole meets the original (and updated) user requirements. As seen in figure 1, summative testing culminates the evaluation process of the user requirements, which have been modified and improved through formative testing.

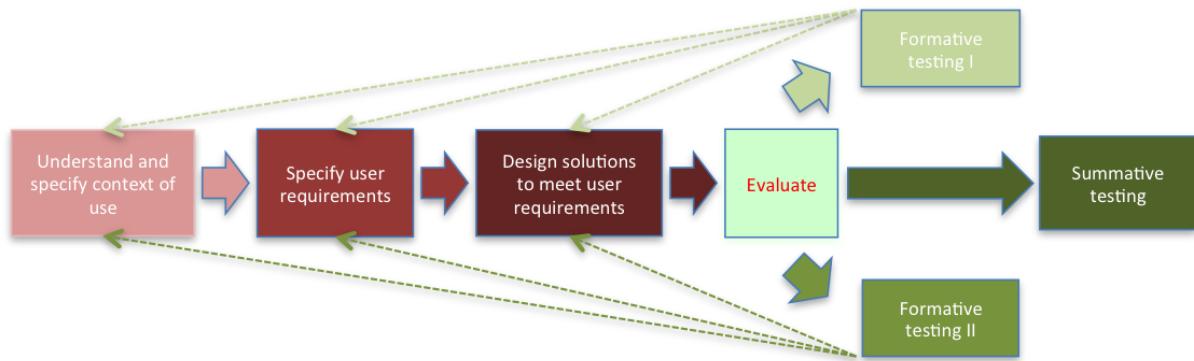


Figure 1: Evaluation process

In this final evaluation round, we have asked the following two questions:

- (1) To which extent does the Final MULTISENSOR System support the user in fulfilling a specific task that is typical for his day-to-day work (task-related evaluation)?
- (2) To which extent does the Final MULTISENSOR System meet the requirements that have been formulated with regard to system usability?

With regard to the first question, the underlying scenarios did depend on the different use cases and will be described in the respective following use case-related sections. With regard to usability, evaluation methodology very much resembles the one that we had chosen for *formative* testing. The main difference is that in this final evaluation round users have evaluated the *integrated* MULTISENSOR system, have assessed how the individual modules *work together* and have tested whether working with MULTISENSOR *in general* is effective, efficient and satisfying. We have also reached out to a larger sample of test users (including the Super User Group - SUG). Although we do not claim that the sample of test users was representative, the results are sufficiently consistent and significant for drawing some clear and authoritative conclusions. Again, these conclusions will be described in the respective use case-related sections.

1.2. Usability Testing at MULTISENSOR

The following text is in some parts identical with Section 1.3 of D8.3 and D8.4. Nevertheless, for better understanding, we have decided to present this very fundamental set of information in this deliverable as well.

1.2.1. General Principles

Usability testing is described as an activity that focuses on observing users working with a product, performing tasks that are real and meaningful to them (Barnum, 2011). More precisely, usability testing needs to measure the level of *effectiveness*, *efficiency*, and *satisfaction* that is experienced by users when they use the MULTISENSOR system in order to

achieve specified goals. ISO 9241-11 (1998) - the relevant DIN standard - provides definitions for these three criteria:

- **Effectiveness:** depends on to which extent the user is able to fulfil the task and to achieve his goals.
- **Efficiency:** depends on how the effort the user needs to invest relates to the accuracy and completeness of the results.
- **Satisfaction:** depends on how satisfied the user is by working with the system.

With regard to the MULTISENSOR evaluation process, we have decided to follow an informal approach to evaluation with real users in a near real-world environment rather than a group of usability experts. The main reason for this decision is that, despite the relevance of the interface design for the project, development has focussed more on specific functionalities that help real users solve problems that are common in their day-to-day work. Also, putting too much emphasis on the interface design would have denied the fact that MULTISENSOR is covering three very different use case scenarios that will ultimately ask for three distinctly different user interfaces.

1.2.2. Summative Usability Testing (Final System)

Despite the different foci of formative and summative testing in general, the evaluation process has been quite similar. Test persons were given specific tasks that they had to perform with the MULTISENSOR system in order to assess its amenities and shortcomings. Similar to the formative testing rounds, we have chosen a mix of *expert reviews* in a *concurrent think aloud process*, followed by a *standard questionnaire* (including some heuristics with regard to the interface) and a *concluding discussion*:

- **Expert reviews:** In the context of MULTISENSOR evaluation, expert reviews means that we have selected specialists from the three different domains (journalism, media monitoring and SME internationalisation), who used the MULTISENSOR system in a typical working environment by performing specific tasks that are common to their day-to-day work.
- **Concurrent think aloud process:** We wanted to understand participants' thoughts when they interact with MULTISENSOR by having them think aloud when performing their tasks. Although this approach interfered from time to time with the work on the tasks itself, it has allowed for more direct and authentic feedback.
- **Standard questionnaire:** After having performed the tasks, participants were asked to fill out questionnaire that enquired about their general experience with the MULTISENSOR system.
- **Concluding discussion:** The evaluation has been concluded by a guided discussion between the evaluator and the participants that allowed for clarifying some ambiguities with regard to the tasks, the system performance and the responses that were given. This discussion was also an opportunity to mention additional aspects that had not been covered by the tasks or the questionnaire.
- **Focus group discussion:** Where possible and appropriate, we have complemented expert interviews by focus group evaluation. In these focus groups, the evaluator presented the

prototype and subsequently allowed participants to test its individual features. This phase was concluded by a group discussion about the benefits of the prototype and its shortcomings.

- **Involvement of the Super User Group (SUG) and other external experts:** Different to the evaluation of the first prototype, this time we have included the SUG and other external partners in the evaluation process. We have also organised a joint workshop and user day with the related FP7 project EUMSSI, to allow for cross-evaluation and thorough assessment of the two projects, as well as their scientific and commercial potential.

The main difference to the previous (formative) evaluation cycles is the number of test persons and a focus on the finished and integrated system.

a) Effectiveness Testing

As mentioned before, ISO 9241-11 (1998) defines usability as *the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*. Effectiveness in this context can be defined as the extent to which the user is able to fulfil a task and to achieve his or her goals. The more accurately the system works, the more effective it is.

We have decided to evaluate the effectiveness of the MULTISENSOR prototype according to the following metrics:

- Number of tasks completed successfully on first attempt;
- Number of persistent errors;
- Number of errors per unit of time;
- Number of users able to successfully complete the task;
- Number of errors made performing specific tasks;
- Number of requests for assistance accomplishing task;
- Quality of output.

b) Efficiency Testing

Efficiency depends on how the effort required to the user needs to complete a task relates to the accuracy and completeness of the results. It is important to understand that efficiency will be judged from a user's point of view. For example, a summarisation tool might be very efficient compared to other automated summarisation approaches, but might not be considered as efficient by the user with regard to the overall task. A journalist, for instance, needs to be sure whether he or she has identified all relevant quotes of a politician with regard to a specific topic, whilst a summarisation algorithm might be considered as efficient from a technical point of view, if its accuracy reaches a level of 85%. In this case the journalist would have to spend time to compare the original text to the summary, making the working process inefficient.

We have decided to evaluate the efficiency of the MULTISENSOR prototype according to the following metrics:

1. Time spent to understand the application and learn about its functionalities;

2. Time spent to perform a particular task;
3. Time spent to perform a task compared to the current method of handling the task;
4. Time spent to perform a task compared to the use of alternative tools.

c) Satisfaction Testing

Satisfaction is defined in ISO 9241-11 (1998) as "freedom from discomfort, and positive attitudes towards the use of the product". Some consider this criterion as even more important than effectiveness or efficiency. If users are pleased with the design of and their interaction with the tool, this *feeling* might even trump the fact that the results of working with the tool are less convincing (Barnum, 2011). As mentioned before, the consortium recognises the relevance of the user interface for the project and the evaluation process. Nevertheless, as the focus will be put on the development of back-end functionalities, the MULTISENSOR evaluation methodology will consider user satisfaction as less crucial than system effectiveness and efficiency.

We have decided to evaluate the satisfaction that a test person experiences when using the MULTISENSOR prototype according to the following metrics:

- Number of users that rate the product as "more satisfying" than their current method of handling the task;
- Number of users that rate the product as "more satisfying" than an alternative tool;
- Number of users who feel "in control" of the product;
- User rating on a five-point scale anchored with "makes me more/less productive";
- Number of users who would recommend it to a friend or colleague.

1.3. Evaluation metrics for theoretical solutions

We analysed the content extraction pipeline (CEP) performance indicators for the different use cases. The goal of the CEP performance analysis is to demonstrate the relation between article length and content to processing time. The processing time can vary significantly, depending on the article length and content. The analysis approach that we followed consisted in choosing 5 randomly articles per use case with the following characteristics: a long article, three of medium length and a short article. The comparison table display results in seconds for every module of the CEP, and we also calculated the average processing time for a set of the articles.

Table 1 presents the results for UC1, in which we notice that the average processing time is 3.36 minutes. Our test has shown that short articles are processed much faster than the articles with longer length so we can conclude that the processing time is relative to the article size. In addition, the processing time increases if the article includes many named entities, pictures or videos.

UC1															
Services	LD	TRANSLA-TION	NER	ENTITY LINKING	ENTITY ALIGMENT	DEPEN-DENCY	RELATION	CONCEPT	SENTI-MENTS	EXT SUMM	CLASSIFI-CATION	CONTEXT	MONGO STORING	TOTAL (s)	TOTAL (m)
Article 1	0	32	7	106.8	1	45	33	4	21	1	17	9	4	280.8	4.68
Article 2	0	21	6	81	0	21	13	2	18	0	18	6	2	188	3.13
Article 3	0	13	2	39	0	4	6	1	10	0	16	1	2	94	1.57
Article 4	0	12	4	34	0	4	5	1	9	0	18	1	1	89	1.48
Article 5	0	29	6	214.2	0	34	27	3	19	0	16	6	3	357.2	5.95
Average	0	21.4	5	95	0.2	21.6	16.8	2.2	15.4	0.2	17	4.6	2.4	201.8	3.36

Table 1: Performance UC1 (in seconds)

Table 2 provide the results for UC2. The same pattern as in the case of UC1 is evident, with an average processing time of 2.33 minutes. The much longer processing times of the entity linking, dependency parsing and relation extraction services affect the overall CEP performance in this regard.

UC1															
Services	LD	TRANSLA-TION	NER	ENTITY LINKING	ENTITY ALIGMENT	DEPENDENCY	RELATION	CONCEPT	SENTI-MENTS	EXT SUMM	CLASSIFI-CATION	CONTEXT	MONGO STORING	TOTAL (s)	TOTAL (m)
Article 1	0	6	2	6	2	6	8	4	12	1	17	3	2	69	1.15
Article 2	0	38	12	90	2	48	39	7	36	1	16	16	4	309	5.15
Article 3	0	29	19	26	1	27	22	3	18	1	16	7	4	173	2.88
Article 4	0	21	4	11	0	8	9	2	12	0	16	4	2	89	1.48
Article 5	0	17	2	8	0	2	4	1	8	0	16	0	1	59	0.98
Average	0	22.2	7.8	28.2	1	18.2	16.4	3.4	17.2	0.6	16.2	6	2.6	139.8	2.33

Table 2: Performance UC2 (in seconds)

Finally, the results for UC3 can be seen in Table 3. The average processing time for articles is approximately 2.26 minutes (similar to UC2).

UC1															
Services	LD	TRANSLA-TION	NER	ENTITY LINKING	ENTITY ALIGMENT	DEPENDENCY	RELATION	CONCEPT	SENTI-MENTS	EXT SUMM	CLASSIFI-CATION	CONTEXT	MONGO STORING	TOTAL (s)	TOTAL (m)
Article 1	0	10	6	4	0	6	8	1	10	0	16	1	1	63	1.05
Article 2	0	30	10	29	1	61.2	45	4	33	1	16	13	3	246.2	4.10
Article 3	0	32	26	23	1	40	31	3	32	1	17	13	3	222	3.7
Article 4	0	16	4	7	0	14	14	2	14	0	16	4	3	94	1.57
Article 5	0	10	1	2	0	4	5	1	9	0	16	1	2	51	0.85
Average	0	19.6	9.4	13	0.4	25.04	20.6	2.2	19.6	0.4	16.2	6.4	2.4	135.24	2.254

Table 3: Performance UC3 (in seconds)

2. Evaluation results

2.1. Pilot Use Case 1: Journalism

2.1.1. Prototype Description and Features

The development of the second MULTISENSOR prototype was based on the list of updated requirements that had been derived from the second evaluation round. These updated requirements have been described in detail in D8.4.

The main requirements for the final development cycle were to improve existing features and functionalities such as summarisation and translation in particular. It also implemented some changes to the MULTISENSOR user interface that were derived from user feedback in the second evaluation. One main aspect of these improvements of the GUI was a simpler display of the results page as we can observe in Figure 2:



Delays to EDF's Hinkley Point C shows UK energy policy must move away from nuclear power

City A.M., GB 16/02/2016

Summary

Profound changes are happening in Somerset, where Britain's first nuclear power plant in 20 years is being built, and in UK energy policy. This means big renewable solar and wind energy, a big grid upgrade, local-scale networks for energy distribution, bridging combined heat and power (CHP) "clean-burn" back-up gas technology, market in [Read the whole summary](#)

Translate into: de ▾

Concepts

investment decision, market price, first nuclear power plant, market value, climate change, coal, carbon, primary energy, energy services, energy supply, UK energy policy, tipping point

Run in-depth semantic analysis [!!](#) Add article to Portfolio ★

Figure 2: Simplified first view of article

With regards to the summaries, users can now choose between a shorter or an extended version (limited to 30% of the length of the original article). These summaries can be translated into German, French and Spanish.

Summary

Profound changes are happening in Somerset, where Britain's first nuclear power plant is being built, and in UK energy policy. This means big renewable solar and wind energy, a big grid upgrade, local-scale networks for energy distribution, bridging combined heat and power (CHP) "clean-burn" back-up gas technology, market in [Read the whole summary](#)

fr
de
es
Translate into: de ▾

Figure 3: Summarisation and Translation

After having assessed the relevance of an article based on summarisation and/or translation, users can now decide whether they want to see more information by initiating the *in-depth semantic analysis*

Run in-depth semantic analysis

Figure 4: Tab for in-depth semantic analysis

Which leads to a new page that displays additional information such as the original article, a list of extracted entities, a tag cloud containing the main concepts as well as a list of related articles.

[◀ Back](#)

2016-02-16 - City A.M. Add to Portfolio

Delays to EDF's Hinkley Point C shows UK energy policy must move away from nuclear power

 Summary
Profound changes are happening in Somerset, where Britain's first nuclear power plant in 20 years is being built, and in UK energy policy. This means big renewable solar and wind energy, a big grid upgrade, local-scale networks for energy distribution, bridging combined heat and power (CHP) "clean-burn" back-up gas technology, market innovation [Read the whole summary](#)

 Translate into: en ▾

Complete article
Profound changes are happening in Somerset, where Britain's first nuclear power plant in 20 years is being built, and in UK energy policy going back to the Blair years, UK energy policy has been long-wedded to the idea that new nuclear will plug the electricity gap and save us from climate change however, it looks economic reality is finally catching up with us.

But EDF's ambitious £18bn construction of Hinkley Point C in Somerset has been beset by problems the company's shares have crashed to half their value a year ago the budget for Hinkley is bigger than EDF's entire market value.

The French firm is selling assets to raise much-needed cash - but there's doubt over whether their "asset" book price matches the market price and their nuclear construction arm, Areva, has been bankrupted by huge costs and time over-runs for the same "brand" of reactor they want to build at Hinkley and now the final investment decision - widely expected to be a mere formality - has been delayed due to EDF's failure to secure the necessary funding.

It looks like the UK is reaching a tipping point, and now it's time to invest in a rational, evidence-based energy policy - before it's too late transitioning to a sustainable UK energy policy won't be easy, and nor will it be cheap however, unlike what's currently in place, it will work

Entities mentioned

 Blair	 Britain
 French	 Hinkley
 EDF	 UK
 European	
 Hinkley Point C	

Extracted Text Concepts 

market value
primary ener

coal
tipping point

climate change
carbon
investment decision

UK energy policy
first nuclear power pla

market price
energy supp

energy services

Figure 5: Semantic analysis results page

Users can also specifically look at articles that include or make reference to multimedia content. Based on all this information, the user can decide whether he wants to add an article to his portfolio in order to run further analysis:



D8.5 – V1.0

[◀ Back](#) [My Portfolio](#)

[Run analysis](#)

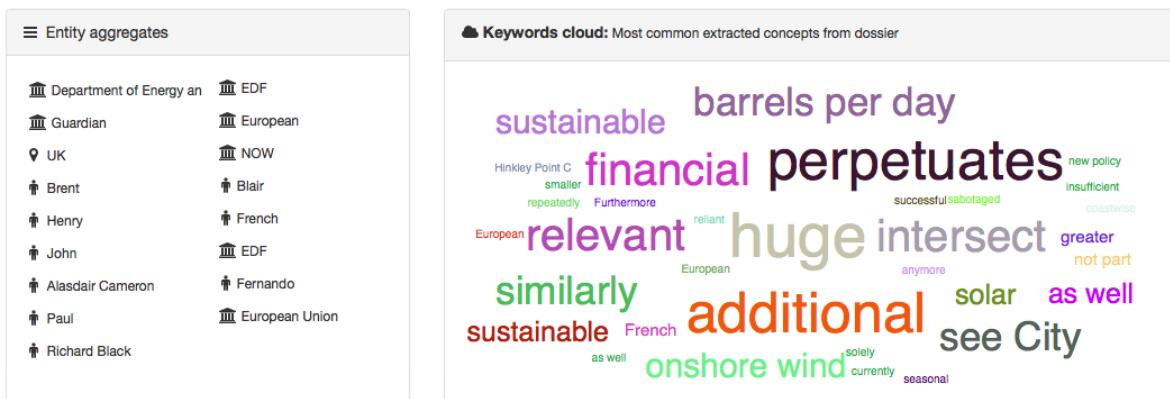
Number of articles: 6

Title	Lang	Source	Date	Category	X
UK energy policy is in disarray - but blackouts are unlikely "Keeping the lights on" is supposed to be the primary duty of energy policy: for good reason. Blackouts are not just difficult for consumers, but dangerous. Our basic infrastructure, from streetlighting to communications and home appliances, is entirely reliant on a dependable electricity supply, a... ↗	en	Guardian	01/03/2016	Economy, Business & Finance	X
Delays to EDF's Hinkley Point C shows UK energy policy must move away from nuclear power Profound changes are happening in Somerset, where Britain's first nuclear power plant in 20 years is being built, and in UK energy policy. Going back to the Blair years, UK energy policy has been long-wedded to the idea that new nuclear will plug the electricity gap and save us from climate change.... ↗	en	City A.M.	16/02/2016	Nature & Environment	X
From airport expansion to energy policy, our politicians are shirking the big decisions – and it's partly our fault It was good to see City A.M. taking aim last month at the absence of long-term planning over energy	en	City A.M.	02/03/2016	Economy, Business & Finance	X

Figure 6: Portfolio view

The Portfolio analysis shows a list of aggregated entities, a cloud of key concepts, a cloud of main topics that are common to all articles in the portfolio and an extended list of related articles:

[◀ Back](#) [My Portfolio: analysis](#)



Topic and Event Detection

Concepts and articles related to your current selection based in topics detected in your dossier

Topic
power
change
plan
energy

Significantly less interest in future UK market, select committee panel says

White House stands by Clean Power Plan

£9 billion a year policy burden could weigh on businesses' ability to deliver jobs and investment – CBI Director-General

#ge16: Which parties have best climate-change policies?

Hamburgs Klima wartet auf Energiewende: Wärme-Dialoo „Kohle-Kerstan“ und immer noch

Figure 7: Portfolio analysis

The combination of all these functionalities enables journalists to analyse individual articles, to assess whether they are sufficiently relevant to be added to their portfolio and to analyse the whole portfolio in order to achieve more complete overview of the chosen articles.

2.1.2. Evaluation Set-Up

As elaborated before, the final *summative* evaluation has focussed on two aspects:

- (1) To which extent does the Final MULTISENSOR System support the user in fulfilling a specific professional task that is typical for his day-to-day work (task-related evaluation)?
- (2) To which extent does the Final MULTISENSOR System meet the requirements that have been formulated with regard to the overall system usability?

Altogether, we involved 35 professionals in the evaluation, with 40% journalists, 37% researchers and 23% participants of other professions.

2.1.3. Task-related evaluation

The test participants were given the specific task to create a portfolio (dossier) consisting of at least five articles relevant to the topic “Energy policy in the UK and the US”. They were asked to explore all available functionalities that were provided by the system and to put particular focus on “summarisation”, “translation” and “in-depth-analysis”. After having completed the portfolio, test persons were asked to run a “portfolio analysis” and to assess its quality. In the following, the description of evaluation results will focus on these four core functionalities. With regard to other modules that have been tested as well (e.g. query) we refer to appendix A.2.

The main question throughout the evaluation was whether a specific feature (i.e. module or functionality) was useful for quickly deciding on the relevance of an article. These feature-related questions were supplemented by questions about the general usability with regard to the MULTISENSOR system as a whole that will be summarised in section 2.1.4.

a) Summarisation

The summaries that the system provides were perceived as particularly useful. Nearly 90% of all test persons agreed or strongly agreed that the summarisation tool was useful for quickly deciding on the relevance of an article. 70% assessed the quality of the summaries as adequate.

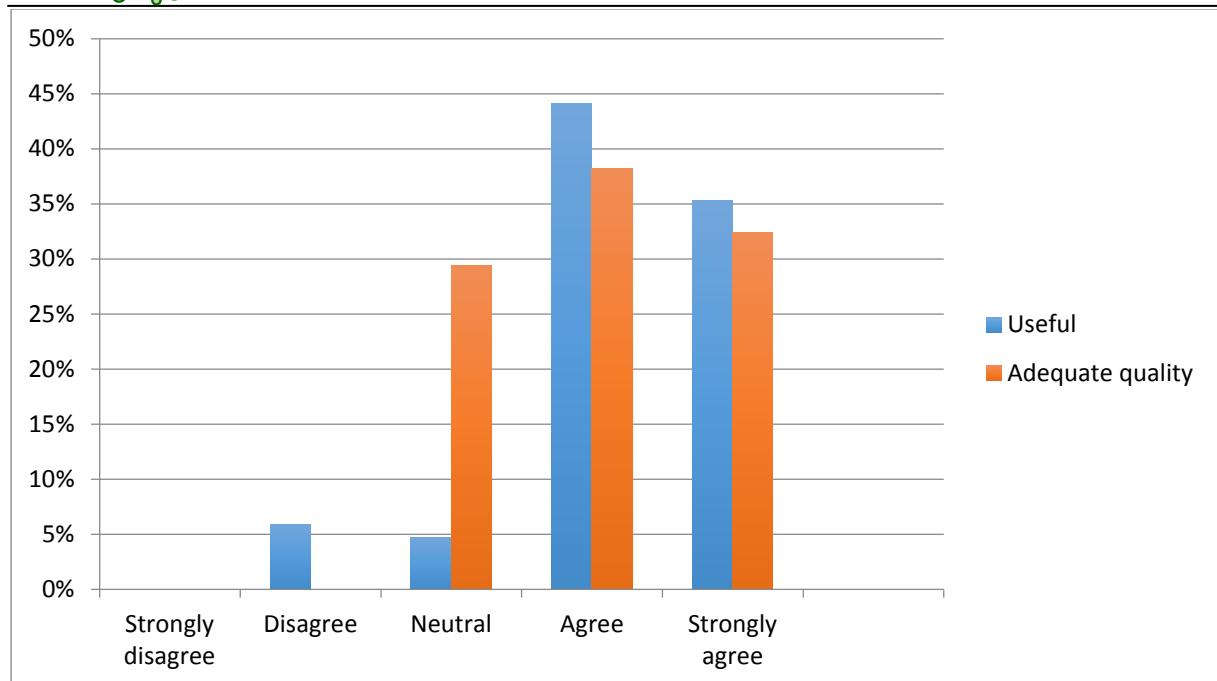


Figure 8: Summarisation Evaluation

b) Translation

The translation module received mostly positive feedback as well. A strong majority of test participants agreed or strongly agreed that the translations were useful for assessing the relevance of an individual article.

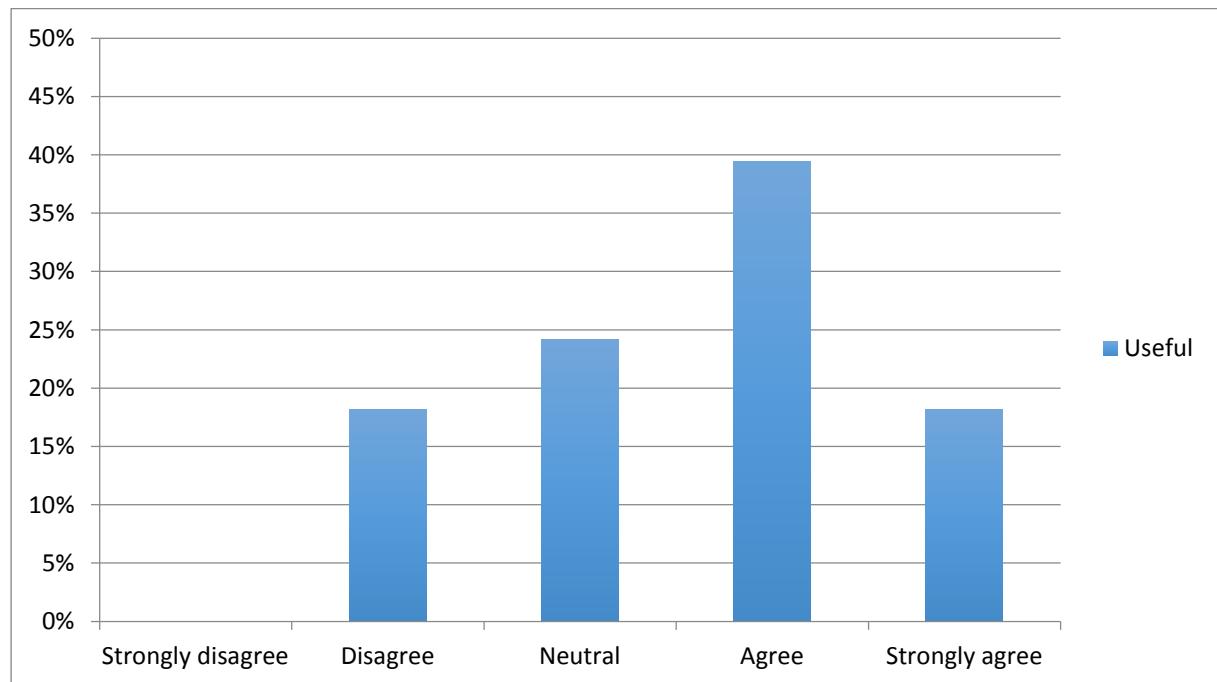


Figure 9: Translation Evaluation

This result was particularly positive as the translations received negative feedback in the first prototype (see the following chart) and were not evaluated in the second iteration at all.

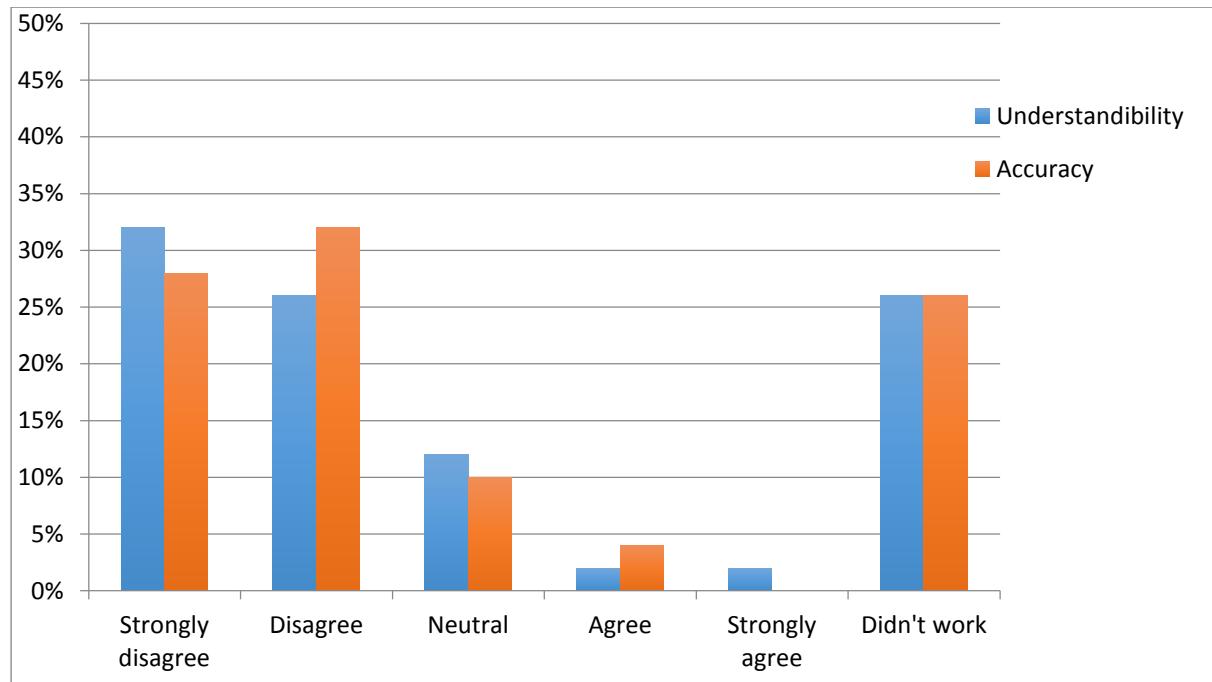


Figure 10: Translation Evaluation 1st Prototype

c) In-depth analysis

The in-depth analysis provided the users with a “list of entities” from the individual article, a “word cloud of key concepts” and a “list of related articles”. We asked the test persons to assess each one of these features, and again the results were mainly positive with a small preference for “related articles”.

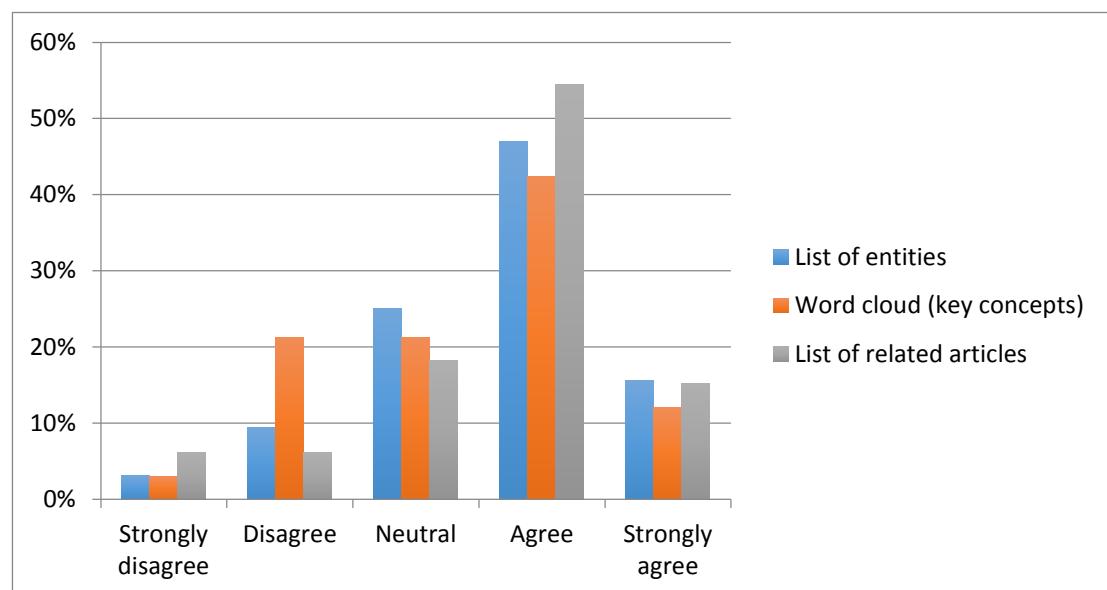


Figure 11: In-depth analysis Evaluation

d) Portfolio analysis

The portfolio analysis provides an aggregated analysis of all the articles that have been identified as relevant and moved to the portfolio by the user. Here, the aim was not about assessing the relevance of an individual article but to achieve an aggregated overview of all selected articles. Evaluation results with regard to the portfolio analysis were a little bit more mixed. More specifically, the aggregated word cloud of key concepts did not convince all users, whilst the list of related articles again was perceived as the most useful one. Altogether, only a minority of test persons disagreed or strongly disagreed with the usefulness of the portfolio analysis in general.

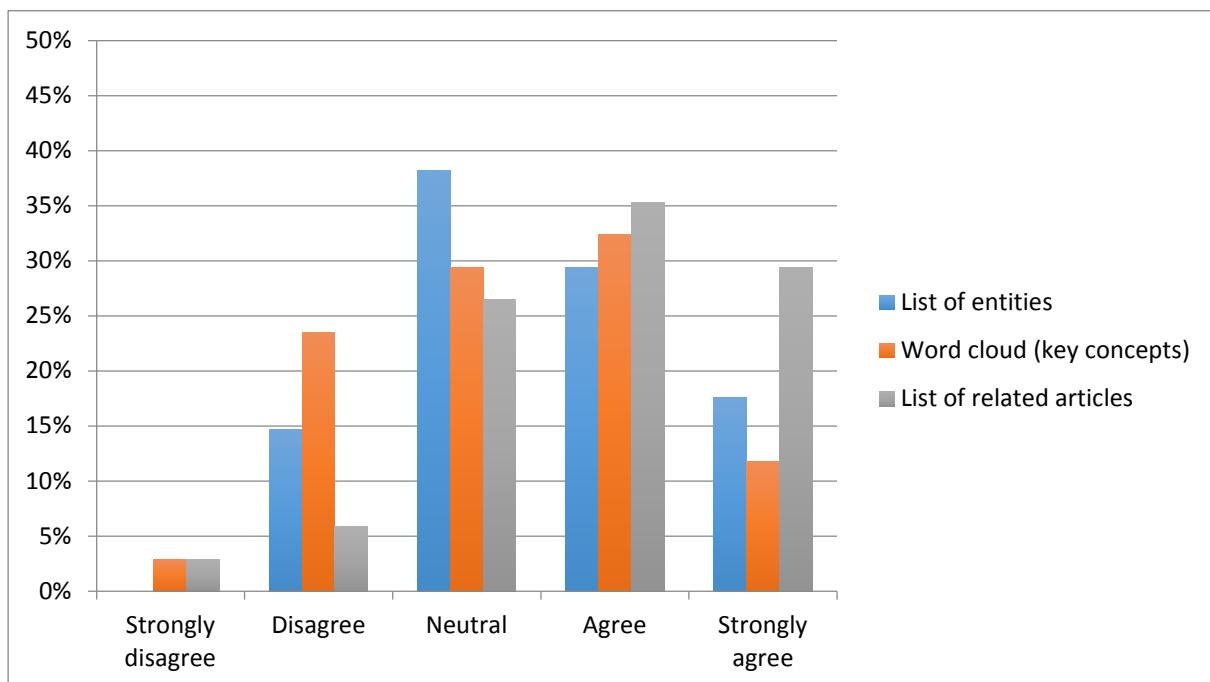


Figure 12: Portfolio analysis Evaluation

2.1.4. Usability testing

In the previous (formative) evaluation iterations, we asked about the effectiveness, efficiency and satisfaction of individual modules (e.g., summarisation). In this final (summative) evaluation round, our aim was rather to assess the effectiveness, efficiency and satisfaction with regard to the integrated MULTISENSOR system and its general performance in supporting a user with a typical task.

a) Effectiveness evaluation

Nearly all test participants were able to successfully complete the tasks that they had been given and perceived the MULTISENSOR system as effective, as shown in Figure 13.

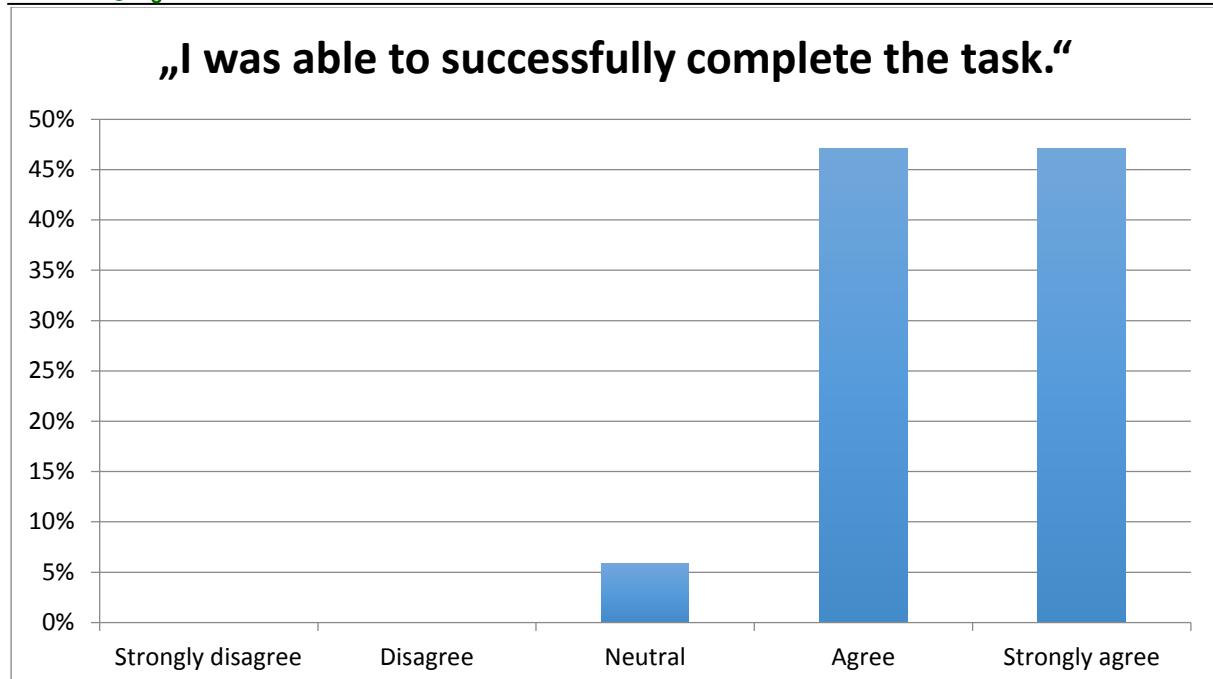


Figure 13: Effectiveness Evaluation

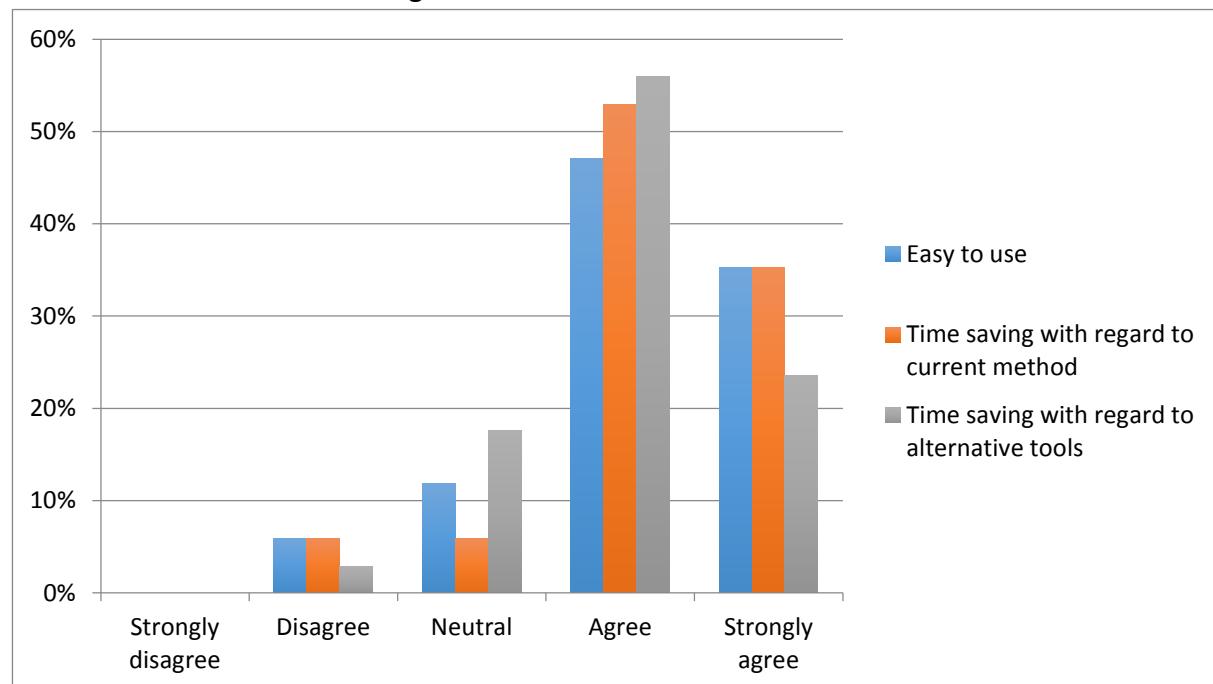


Figure 14: Efficiency Evaluation

b) Efficiency Evaluation

Having assessed the effectiveness of the MULTISENSOR prototype, participants were asked to evaluate its efficiency. Efficiency depends on how the effort the user needs to invest relates to the accuracy and completeness of the results. We asked how easy the prototype was to use and on how much time it took to perform the tasks. Generally, the results have

confirmed the very positive outcome that we had already experienced in the first two evaluation rounds, as shown in Figure 14.

c) Satisfaction Evaluation

More than 75% of all test participants perceived the interface as intuitive and assessed the use of MULTISENSOR as an overall satisfying experience. In addition, a clear majority said that they felt in control (67%) and more productive (62%) when using MULTISENSOR. A further and even 70% would recommend the system to others.

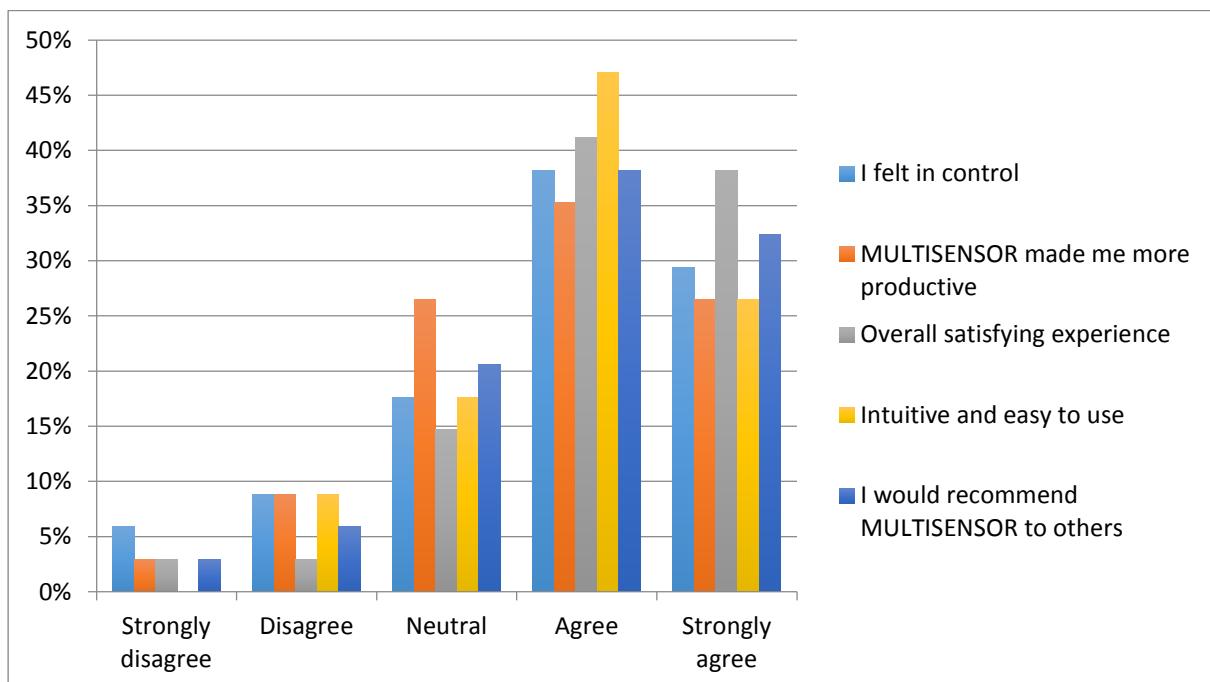


Figure 15: Satisfaction Evaluation

2.1.5. General Comments

Test persons made a number of very diverse comments ranging from detailed feedback on individual modules to suggestions for how to improve the user interface. These comments can be found in the Evaluation Summary (appendix A.2). We also asked test persons to tell us which functionality of the MULTISENSOR system they perceived as “most promising and suitable” for further development and subsequent exploitation. Most of the tested modules were mentioned at least once, but “summarisation” was clearly the functionality that got most of the votes.

2.1.6. Evaluation Analysis and Final Assessment

Overall, evaluation in the journalistic use case has shown very positive results. With regard to nearly all individual functionalities that have been tested, a clear majority of up to 75% of test persons agreed or strongly agreed with the usefulness of the respective functionality. But also in the rare cases when this majority was not achieved (e.g. portfolio analysis), only a minority of less than 30% considered the respective functionality as not useful.

Nevertheless, these results need to be put into context. We did not ask to assess the quality of each individual functionality or module for itself. Instead, the leading question in this evaluation was whether MULTISENSOR is useful for fulfilling a very specific professional task. The testing showed that the integration of different functionalities and modules is mostly perceived as useful. But it would not be legitimate to conclude that the development of individual modules has reached a level of quality that would allow for immediate exploitation in the market. It should in any case be obvious that it will take some time and effort in general until automatic summarisation or machine translation have reached a quality level that is comparable to the quality of human work. But also the fact that not all functionalities were assessed as equally useful shows that a very positive overall evaluation does not imply immediate exploitability.

However, the potential is apparent. When developing the user requirements at the beginning of the MULTISENSOR project, we identified a possible strong USP for MULTISENSOR from a journalistic point of view (see D8.2, page 20):

Automatic summarisation of heterogeneous and multilingual digital information in English.

The MULTISENSOR summarisation tool did not only receive good feedback with regard to its performance but was also considered as the most promising and suitable functionality for further development and exploitation. This confirms the original hypothesis and shows at the same time that the consortium succeeded in developing a system with real exploitation potential. D9.7 will elaborate how the consortium intends to utilise this potential in the future.

2.2. Pilot Use Case 2: Commercial Media Monitoring

2.2.1. Prototype Description and Features

After the evaluation of the 2nd MULTISENSOR prototype, development focused on the one hand on streamlining already existing features in order to achieve a smoother workflow, and on the other hand on making usability improvements as suggested by the evaluators. Several requested features were integrated for the first time and newly available.

The MULTISENSOR PUC2 final prototype is divided into four different areas (Figure 16):

- 1) a search area, where queries can be performed on the data in the MULTISENSOR news repository – this area offers tools that support data curation and speed up the selection process;
- 2) an analysis area, where the previously selected content is analysed and visualised;
- 3) an influencer area, where the user can search for influencers and networks/communities from the household appliances domain; and
- 4) a profile area, where the user can configure and update his or her ongoing search or analysis projects.



D8.5 – V1.0



UC2 ANALYSIS EVALUATION UC2 EVALUATION ENGLISH
SEARCH ANALYSIS INFLUENCERS PROFILE

Figure 16: PUC2: Top Menu.

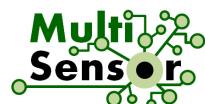
Having logged in and selected a pre-existing or new profile, the user can open the search tab and enter a search string. New to the final prototype is a semantic search, meaning e.g. that entering a term like “energy consumption” will deliver multi-language results as long as no restrictive filters have been set. The interface itself has been debugged but is otherwise unchanged when compared to the 2nd prototype.

Search results are returned in a single-article list (optionally also clustered), from which the user can easily select relevant content. Each article shows metadata such as sentiment and category besides source, date, language and country (Figure 17).

The screenshot shows a single article summary. At the top, there's a header with the title "Sicherheitsbedenken im Smart Home" and a link. Below it, metadata includes "Source: Mobile Business", "Date: 2016-03-10 12:36:00", "Language: DE", "Country: DE". A sentiment bar shows a red-to-green gradient with a value of 3/8. The category is listed as "Science & Technology". The main content discusses security concerns in smart homes. At the bottom, there are several buttons: "show Summary" (with a pencil icon), "show keyword-based summary" (with a pencil icon), "show Translation" (with a flag icon), "hide Entities" (with a video camera icon), and "show full text" (with a dropdown arrow icon). Below these buttons is a section titled "Entities:" containing a grid of icons representing various entities like French, Kit, Markus, Matic, More, Norbert, Reza, Siri, Smart, So, Thierry, Thomas, U, Samsung, Google, Miele, RWE, Deutsche Telekom, Deutsche Telekom AG, Cisco Systems, Cisco Systems, Inc., EnBW, ABB, Apple, Bosch, Eco, Ocean, and Amazon.

Figure 17: PUC2: Single Article View with Entities Displayed.

Via buttons, the user can select additional information that may help him to assess the relevance of the displayed content faster (Figure 17). Available are *summarisation*, *translation*, *detected entities* and *full text*. As new feature, keyword-based summarisation has been integrated (Figure 18). This new functionality allows the user to select a keyword of his choice, and the summary generated in the following will put special emphasis on this keyword. Detected entities are offered for selection to speed up the process.



D8.5 – V1.0

Select from detected entities:

Smart Smith Panasonic Panasonic Corporation Siemens
Hitachi Hitachi, Ltd. Mitsubishi Mitsubishi Group Siemens AG
AWE Bosch

Select active search term(s):

energy consumption

Enter custom keyword(s):

+
Close Go

Figure 18: PUC2: Keyword-Based Summarisation, Selection Screen.

The value of this functionality is that media monitoring is usually booked by clients with a particular interest in their company's part in the media coverage and want their summaries to contain information related to it.

After selecting and storing all relevant content to the profile, the user can evaluate the collected media coverage in the analysis area (Figure 19, 20).

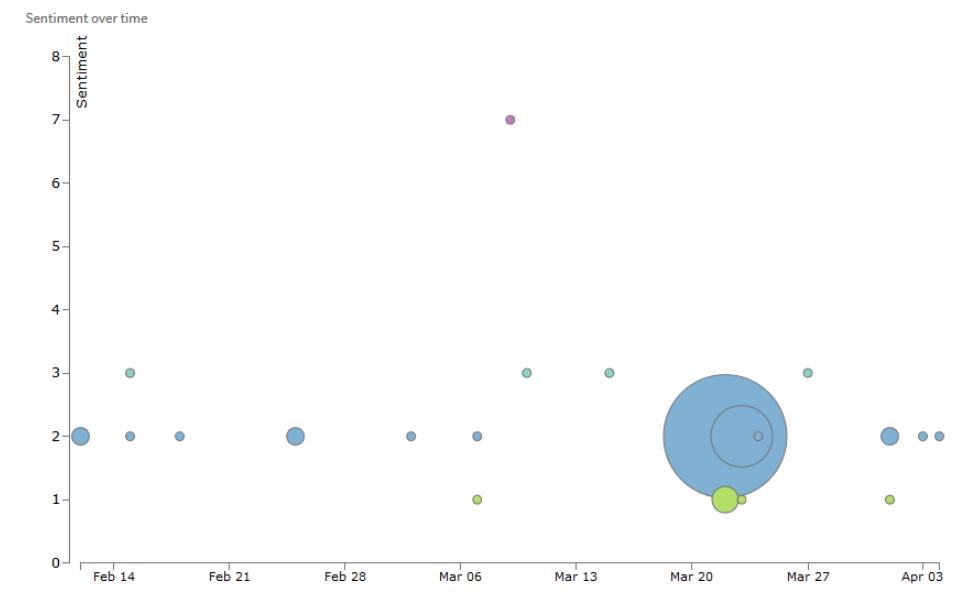


Figure 19: PUC2: Analysis Charts (1).

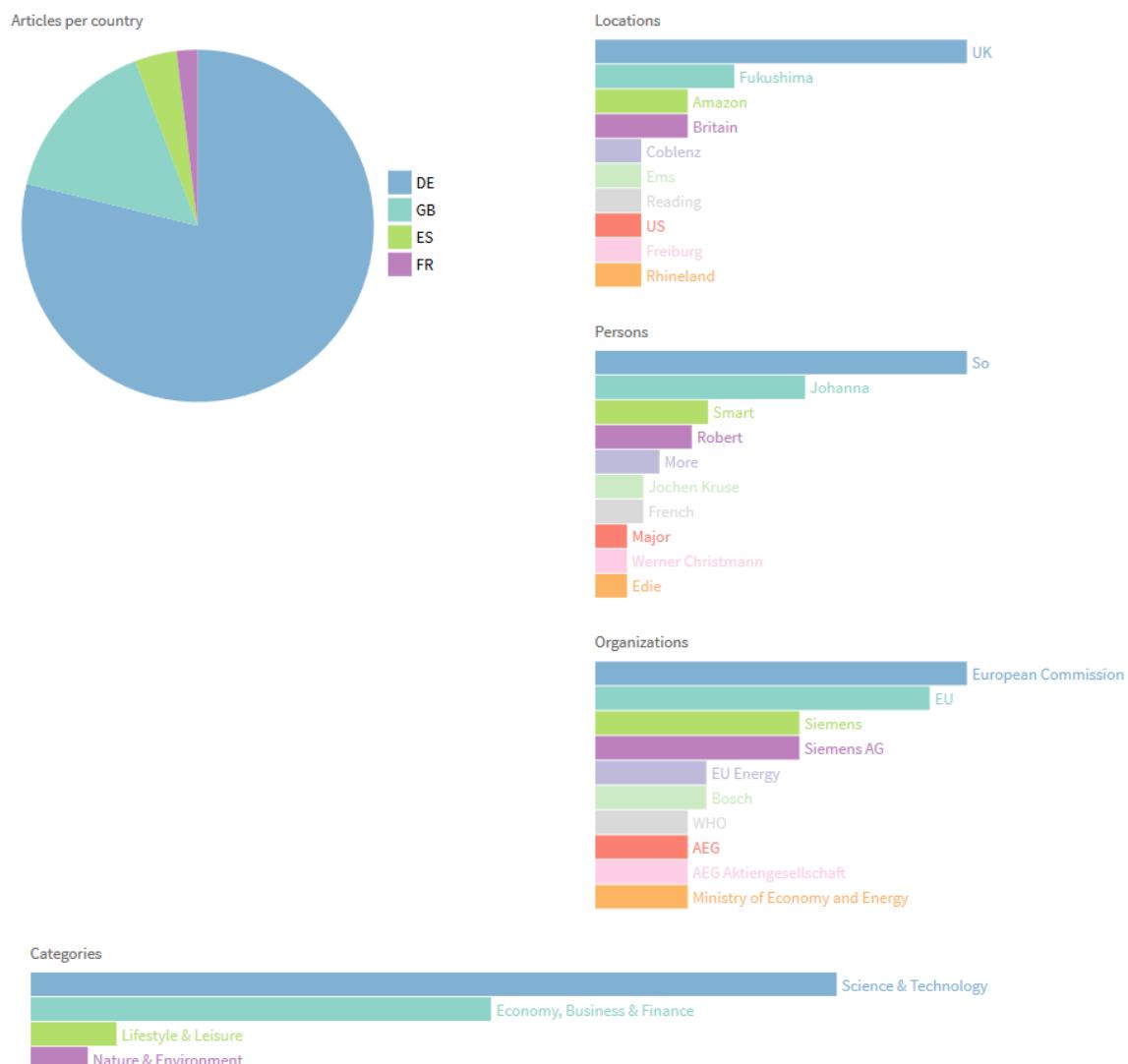


Figure 20: PUC2: Analysis Charts (2).

The user can click on the charts and see the relevant content behind the bars in a drill-down feature to the individual articles. As new feature, a multi-document summary (Figure 21) is created and displayed through the same click.

Summaries of all selected articles (4) [hide](#)

Nachdem der Konzern 2007 noch das teuerste Unternehmen Frankreichs und nach Shell der zweitteuerste Konzern Europas war, wurden seine Aktien im vergangenen Dezember von einer kleinen Immobilien gesellschaft aus dem Börsenindex CAC-40 gedrängt.

Brüchige Grundlage EdF ist, anders als Areva, zwar nicht in der Existenz gefährdet, doch der Stromproduzent steht ebenso vor grundlegenden Fragen.

Beide Sondermerkmale der französischen Energiepolitik – der besonders hohe Nuklearanteil von 75 Prozent der Stromproduktion sowie ihre Umsetzung durch große Staatsunternehmen – geraten an ihre Finanzierungsgrenzen.

Schwere Managementfehler wie der Kauf einer wertlosen Uranmine in Namibia sowie die Selbstüberschätzung beim Bau eines finnischen Kernkraftwerks haben fast zum Zusammenbruch geführt.

Im finnischen Olkiluoto und im französischen Flamanville treten daher seit Jahren erschreckende Bauverzögerungen und Budgetüberschreitungen auf.

Zu lange weggesehen Die staatlichen Verwaltungsräte sahen jahrelang weg oder ließen sich blenden, wenn die langjährige Vorstandsvorsitzende Anne Lauvergeon („Atomic Anne“) in ihrer energischen Art die Pläne vorstellte.

Die längst abgeschriebenen 58 Nuklearreaktoren des Landes stellen jedenfalls kostengünstig Elektrizität her.

An anderer Stelle hat die öffentliche Hand wiederum Mittel blockiert, wenn sinnvolle Investitionen anstanden.

Strompreiserhöhungen blockierten sie, um die Kaufkraft der Franzosen zu schonen.

Die französische Energie wirtschaft ist zu Recht stolz darauf, weitgehend CO₂-frei zu sein.

Figure 21: PUC2: Multi-Document Summary.

The influencer section (Figure 22) shows twitter content from the household appliances domain. The user can rank the influencers according to several metrics: the MULTISENSOR influence score, the number of tweets, followers and number of people following are available.

Influencer meta data

The influencer score indicates how influential a user is in a social network, which relates to the number of his/her followers and how often these followers share that user's content. The higher the influencer score is the more influential is the user.

Image	Username	Name	Influence	Tweets	Following	Followers
	Tip_Lina780	Lina	97	20362	339	511
	95Gardner	Rosalinda Gardner	88	19022	353	524
	auld1976	Auld Pearl	87	18142	322	478
	delcie_ms	Delcie Ahumada	82	20623	371	560

Figure 22: PUC2: Influencer Meta Data.

Also, the user can evaluate communities (Figure 23) active that day through an interactive network chart.

Communities by mentioning

Dots represent twitter users, users are linked through mentioning. The more a user is mentioned by other users, the more edges target this specific user. Dot sizes are calculated accordingly.

Note: this chart only shows communities of the current day. It tends to grow a little during the day.

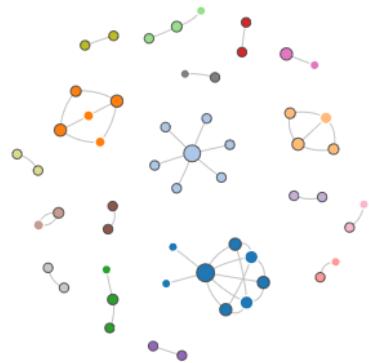


Figure 23: PUC2: Community Network Chart.

2.2.2. Evaluation Set-Up

As already laid out for Pilot Use Case 1, we focused on summative evaluation in this final evaluation round, which means that the usefulness and usability of the services and tools delivered by MULTISENSOR in general stood in focus. Less emphasis was placed on the qualitative aspects of the displayed results.

Since the exploitation plans for the media monitoring use case foresee a mostly modular exploitation of individual features, we understood the final MULTISENSOR PUC2 prototype as a whole to be a demonstrator for the implemented technologies and workflows rather than a stand-alone platform. As argued in earlier deliverables, players in the media monitoring market tend to have rather complex and intricate production processes and are unlikely to exchange their production system for another readily. Modular integration of individual services has a much higher chance of being a marketable approach than trying to establish a new stand-alone system.

The evaluation scenario reflects this modular thought and does not put emphasis on evaluating a complete and all-encompassing workflow. Rather, we were interested in seeing how the results MULTISENSOR delivered would be accepted at the different simulated points of a near-real-life workflow.

Just like before, the evaluation was conducted through think-aloud interviews as well as remote evaluation sessions and hands-on evaluation during the 2nd Open User Day in Barcelona in September 2016. In total, this third evaluation round consisted of 19 participants, out of which 63% had a media monitoring background and the remaining 37% of participants were users with a general interest in the media monitoring results as potentially delivered by MULTISENSOR. Some of the evaluators were members of the MULTISENSOR User Group.

2.2.3. Task-Related Evaluation

The users were presented with a near-real-life set of tasks that depict important steps in the daily working-routine of a media monitoring employee. Testing users were asked to make a free query in the search section with search terms from the household appliances domain, for example to search for “energy consumption” or a specific appliances brand in the repository. In the following, users were required to select a number of relevant articles for an imagined household appliances customer while assessing the relevance of the articles using the tools and services provided by the MULTISENSOR interface, such as summarisation, translation, sentiment etc. Having completed their selection, the users were asked to move on to the analysis section and assess the usefulness both of the displayed charts and the multi-document summarisation tool behind them. In a third part, the evaluators switched to the influencer section in order to look for the most important household appliances influencers and networks.

The complete questionnaire for PUC2 evaluation can be viewed in appendix B.1.

a) Search Section / Single Results List

Though not evaluated by a dedicated question in the questionnaire, the semantic search with its multi-lingual results received praise from several evaluators:

“The semantic search is very good and would be a great help!”

“Semantic search is impressive.”

The complete user comments can be viewed in appendices B.2 and B.3.

When evaluating the provided features for faster data curation, the summarisation stood out with good results just as in the previous evaluation rounds. The feature received 74% of affirmative answers when asked about its usefulness; 68% of the users attested the extractive summarisation to have adequate quality to speed up the article selection process.

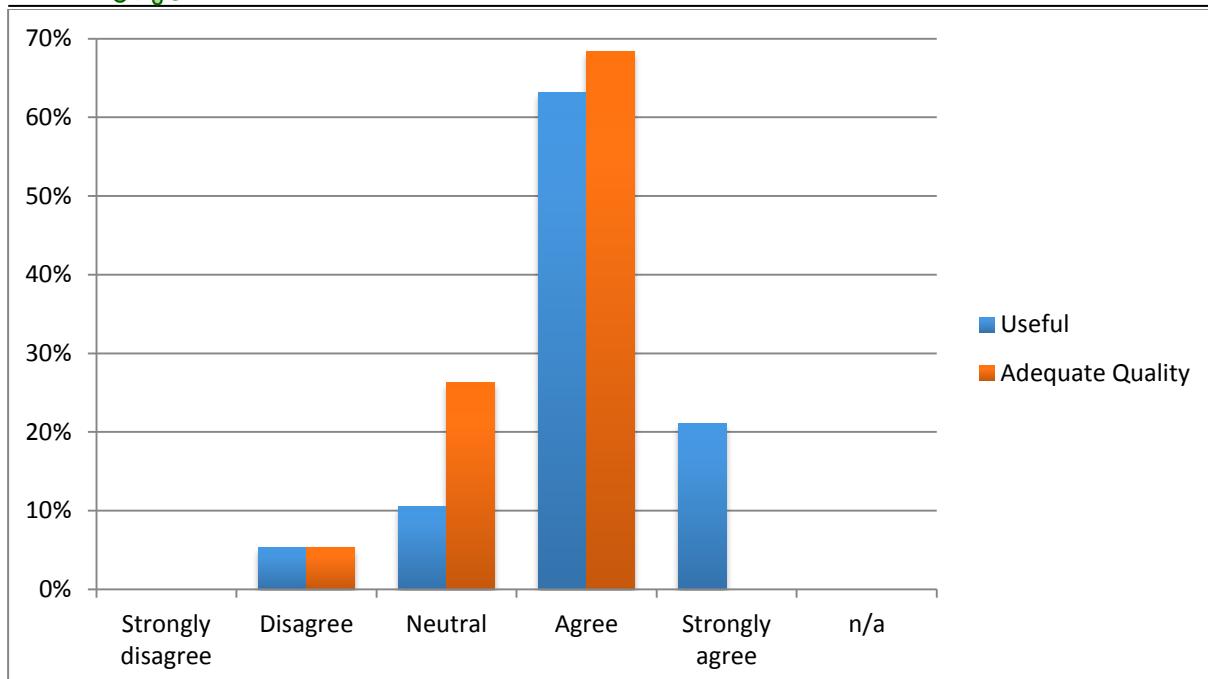


Figure 24: PUC2: Extractive Summarisation.

When compared to extractive summarisation (Figure 24), the new feature offering keyword-based summarisation (Figure 25) received slightly weaker feedback, with nevertheless 60% of participants agreed or strongly agreed that the resulting summary adequately mirrored the content of the article from the client's point of view. Since this feature was only recently integrated and tested for the first time, slight usability issues in the interface may have lowered feedback for the keyword-based summarisation.

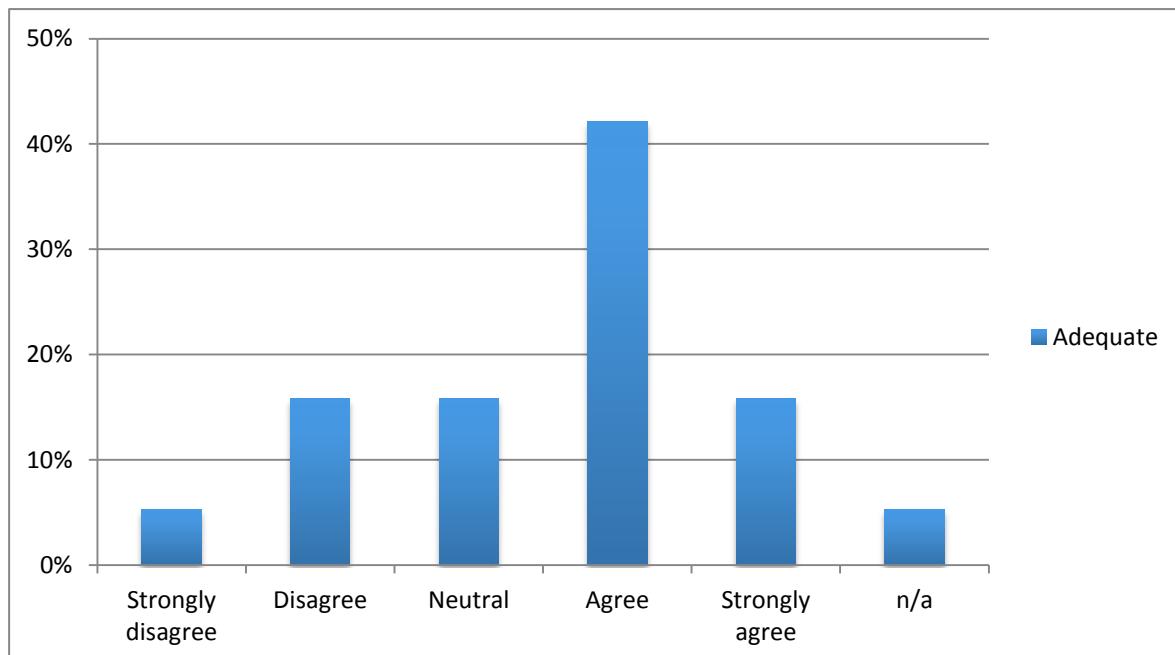


Figure 25: PUC2: Keyword-Based Summarisation.

Evaluation feedback for the translation feature (Figure 26) has improved compared with previous evaluations, and 75% of the users found the integrated translation to be a useful tool that helps to quickly assess the relevance of an article for a media monitoring client and get a first grasp of its content.

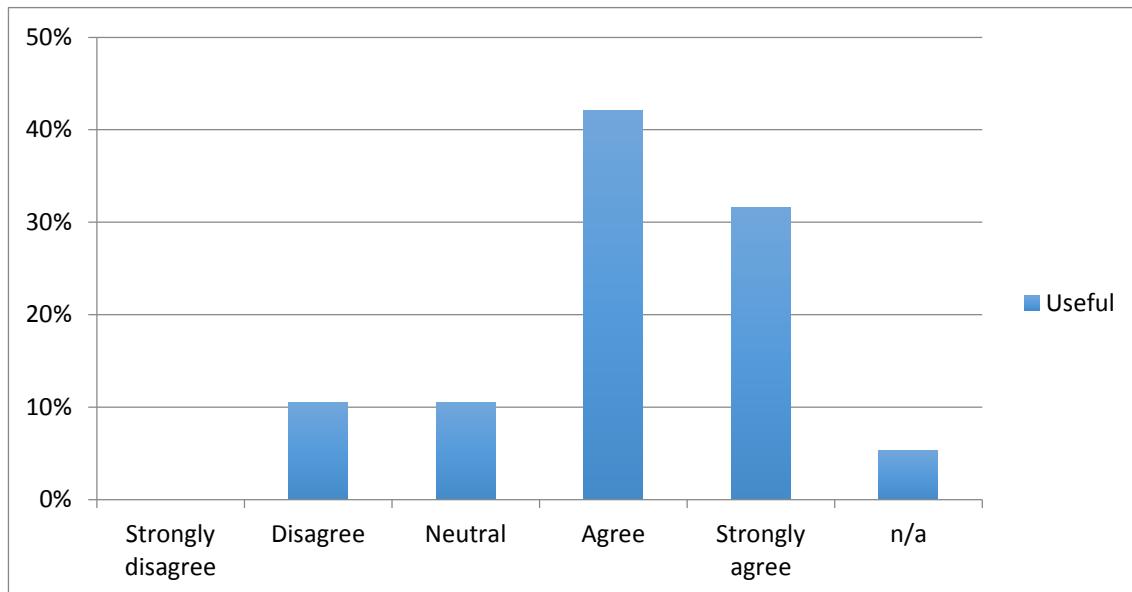


Figure 26: PUC2: Translation.

For the three context features in the article results list, namely the display of the detected entities, categories and sentiment, we asked the users if these were useful for quickly selecting relevant content.

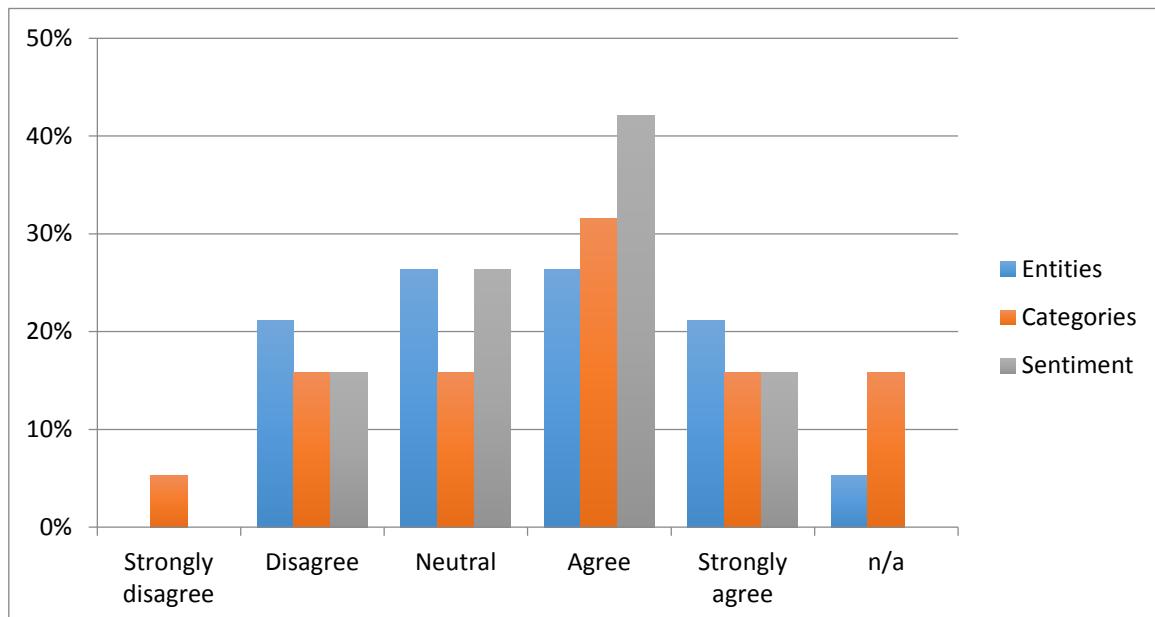


Figure 27: PUC2: Context Information.

While their relevance in a broader media monitoring work context had been affirmed in the previous evaluation rounds, all features received mixed results now. 58% of the participants determined the sentiment to be useful for quickly deciding on the relevance of an article, while feedback on entities and categories scored with 47% and 48% on the positive side. All three features were evaluated a second time in the analysis section, where the analysis charts provided more condensed information for the set of articles selected by the client.

b) Analysis Section

Having completed their article selection, the evaluators were asked to go to the analysis section of the PUC2 interface and assess the displayed content. We asked them if the charts in the analysis section provided helpful information about the article selection. The charts on display showed the extracted named entities, split per persons, companies/organisations and locations, as well as a chart with the article categories and the sentiment of the articles over time.

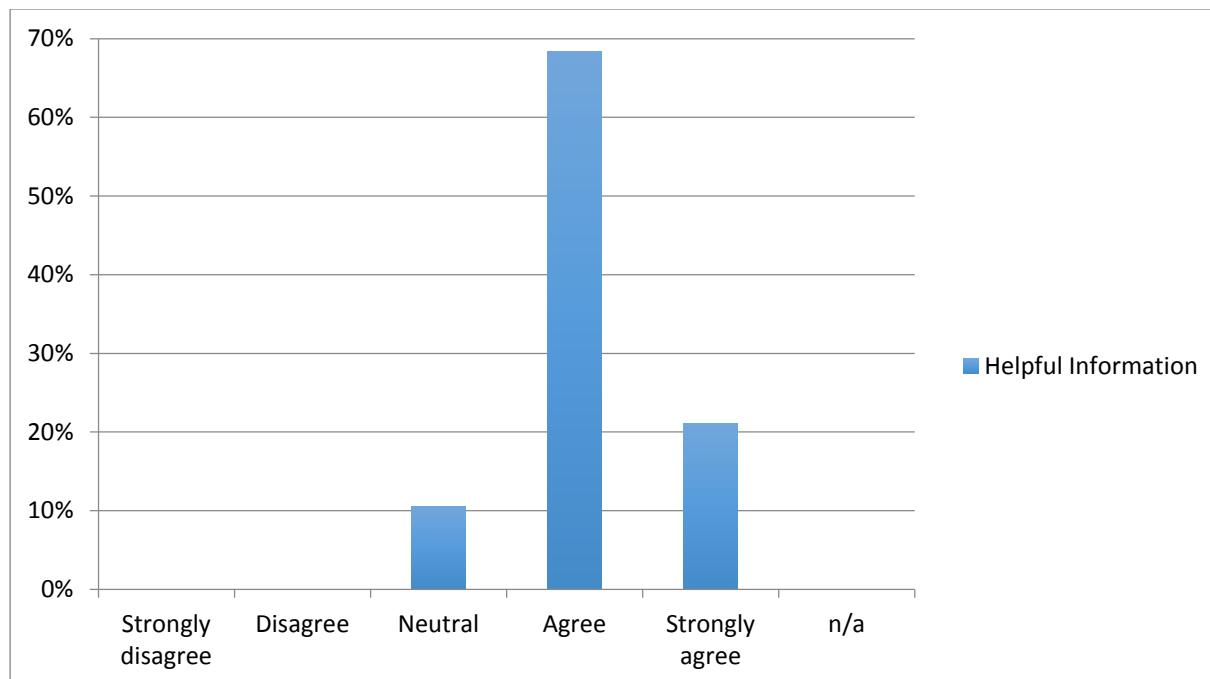


Figure 28: PUC2: Analysis Charts.

89% of participants found the displayed content to be helpful and relevant (Figure 28). Comparing these numbers with the lower results for the same extracted information in the single results list, the chart evaluation result suggests that the users perceive a higher value in automatic dashboard creation than in displaying the extracted information in the data curation process.

The users were asked to click on the interactive analysis charts in order to see the drill-down effect and to create a multi-document summary. The latter feature had only recently been integrated and was the first time in a user test – a fact that is mirrored in the evaluation results. While 52% of the users agreed that a feature like this would be useful in an analysis context, we received a lot of feedback with hints for improvement and new requirements

that users would like to have for this kind of feature. All user comments can be viewed in appendices B.2 and B.3. However, here we present the most relevant on the possible improvements on multi-document summary.

"This is a nice idea for analysts. However, presently the quality is not good enough. It seems that the system only chooses the first few sentences of every article. I would prefer to have a list display instead of a text block, as the relations between sentences become unclear. This list should contain e.g. the sentences with the highest sentimentality or the most relevant statement otherwise. It should be ranked and offer the possibility to deselect irrelevant information."

"I would not read the multi-doc summary in this format. Instead, I would like it to focus on similarities and differences between the articles and I would prefer a list of bullet points."

"Basically, this is a good idea, but the summary needs more structure. It would be good to focus on the main statements of the articles and always mention the source of the information. I would like to see contrasting and overlapping information visualised. Quotes might also be interesting. Like in the keyword-based summarisation, you should probably have a focus on your client's interests."

31% of the evaluators denied that the multi-document feature would be of use to them in the current form, mostly because of unstructured display, but also because of different wishes for the content of the multi-document summary (Figure 29). Due to this user feedback and in order to increase acceptance, the summary tool was adapted after the evaluation to display paragraphs rather than a text block.

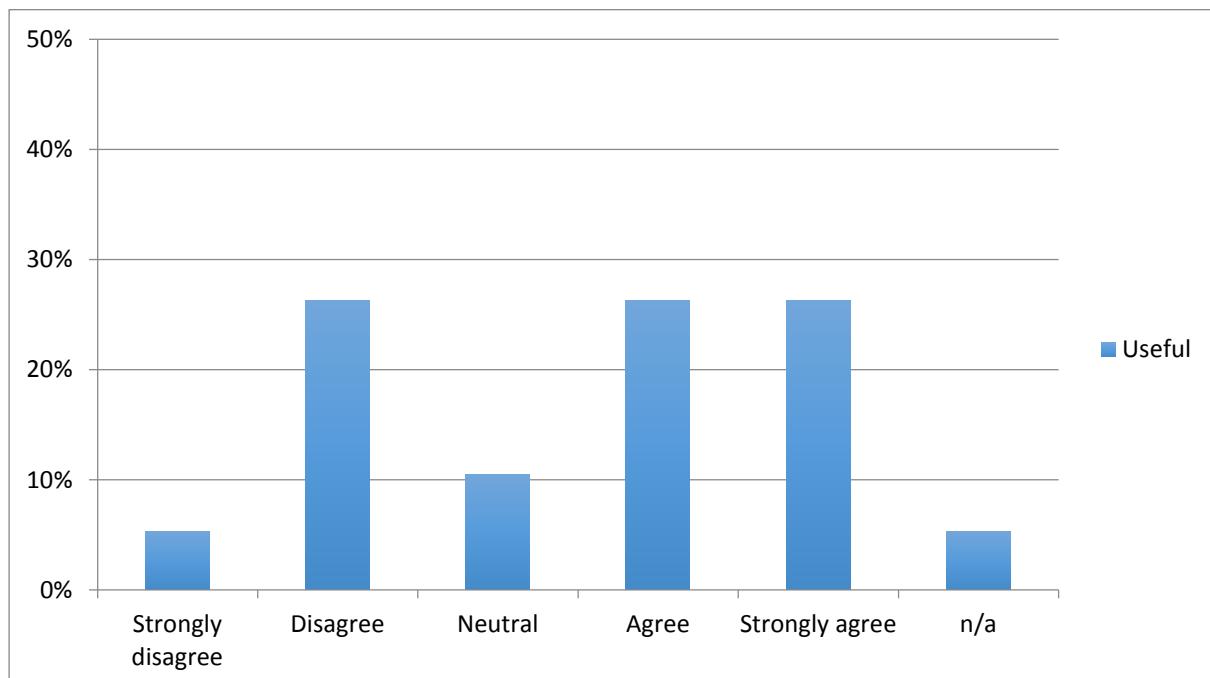


Figure 29: PUC2: Multi-Document Summarisation.

Summed up, it can be said that in order to make multi-document summarisation usable in a product context, more emphasis needs to be placed on a user-friendly display of the resulting information, taking into account the particular information interests of the readers. The widely positive reactions to the idea of having automated multi-document summarisation within a tool is however encouraging to proceed with the development of this idea.

c) Influencer Section

In the third part of the evaluation session, the users were directed to the Influencer Section of the PUC2 interface. In here, they were asked to identify the most important influencers for the household appliances use case using the MULTISENSOR influence score and other established metrics displayed along with them. A very high number of 89% of testers found the influential user information to be a useful aid for this task (Figure 30). Nevertheless, most of the users also expressed some general scepticism toward a new score and wanted to understand precisely how it is calculated. Almost all of them criticised the scale of the influence score to be too small, which led us to change improve this display directly after the evaluation.

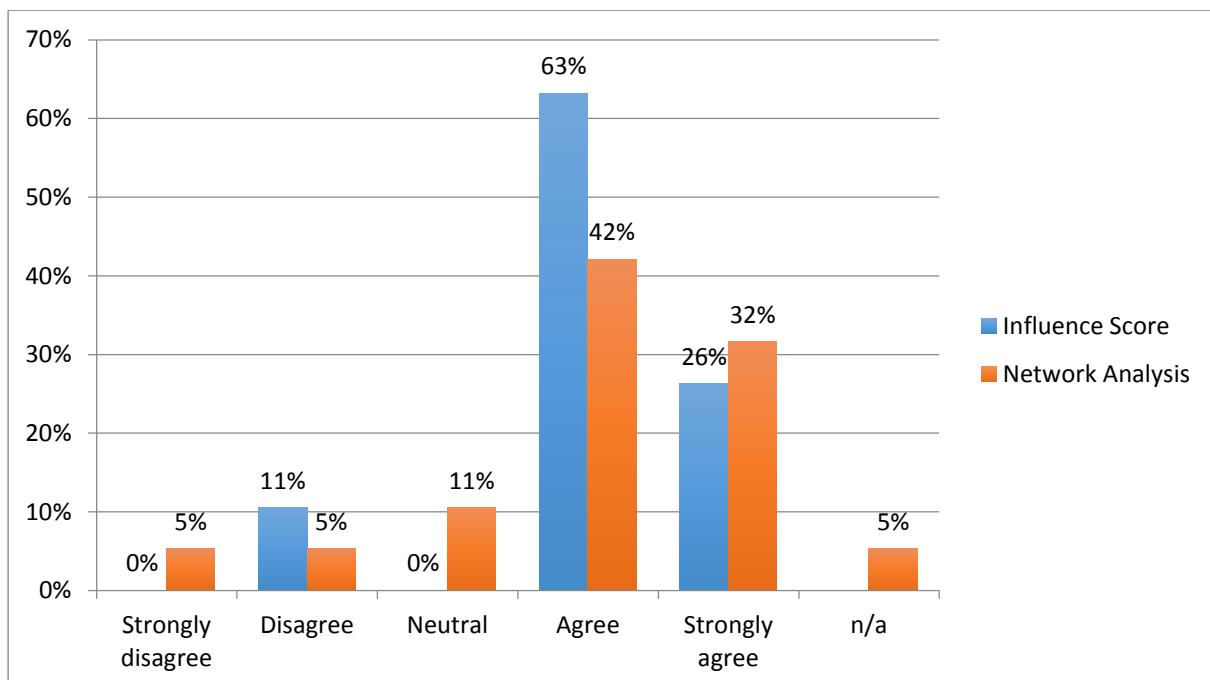


Figure 30: PUC2: Influencer Information.

Some input was also given for the improvement of usability:

"When I click on an influencer, I cannot immediately see why he is considered influential. The relevant post should be directly linked or be displayed otherwise."

Users also emphasised that to them, influence was strongly connected to a topic of interest:

"Generally, I would want to restrict the displayed followers to topics that interest me, e.g. coffee machines or the like so I can reach out to the most interested influencers."

"To me, an influence score should not only consider the followers and share, but also how relevant an influencer is with regard to the topic I am interested in, e.g. fashion. Making a qualified statement who is influential for a topic is a very difficult decision. Maybe looking at the profile descriptions could help, or setting a threshold number of relevant posts that needs to be surpassed before being considered an influencer."

Overall, the received feedback on the influence score suggests that in the media monitoring context, the influence score can be an asset, when the calculation basis is transparently explained and the topical relevance of the shown influencers is not only ensured by the system but becomes also obvious to the user, e.g. by linking or displaying relevant content created by the influencer.

The network analysis, on the other hand, was positively taken up with 74% of the users finding it helpful in detecting relevant communities. Just as for the other features, relevant input was given with regard to usability improvements and potential new requirements for future development, e.g. calling for more user metadata in the mouse-over.

2.2.4. Usability Testing

As explained above and also in consistency with the previous evaluation rounds, the near-real-life evaluation tasks were followed by questions targeting the MULTISENSOR system as a whole – this time with a summative focus. The less-than-complete media monitoring workflow that the MULTISENSOR prototype offers has led to a modular exploitation approach for PUC 2, and naturally, the fact that only a part of the process can be adequately depicted is mirrored also in the evaluation results for the MULTISENSOR system as a whole. Namely, comparability with other known tools has proved to be difficult to assess. The same goes to some degree for the assessment of efficiency and satisfaction. Nevertheless, as evaluation of effectiveness, efficiency and satisfaction have been a central part also of the previous two evaluation rounds, the results provide good insight.

a) Effectiveness Evaluation

79% of the evaluators agreed or strongly agreed they were able to successfully complete the tasks, 58% of them at first attempt and without assistance (Figure 31). Errors were reported by 21% of the users, mostly due to a usability issue in the keyword-based summarisation tool that has been fixed in the meantime. These results are largely corresponding to the answers from the evaluation of the 2nd prototype. Due to the fact that a large part of the features was implemented and tested for the first time, this is not surprising.

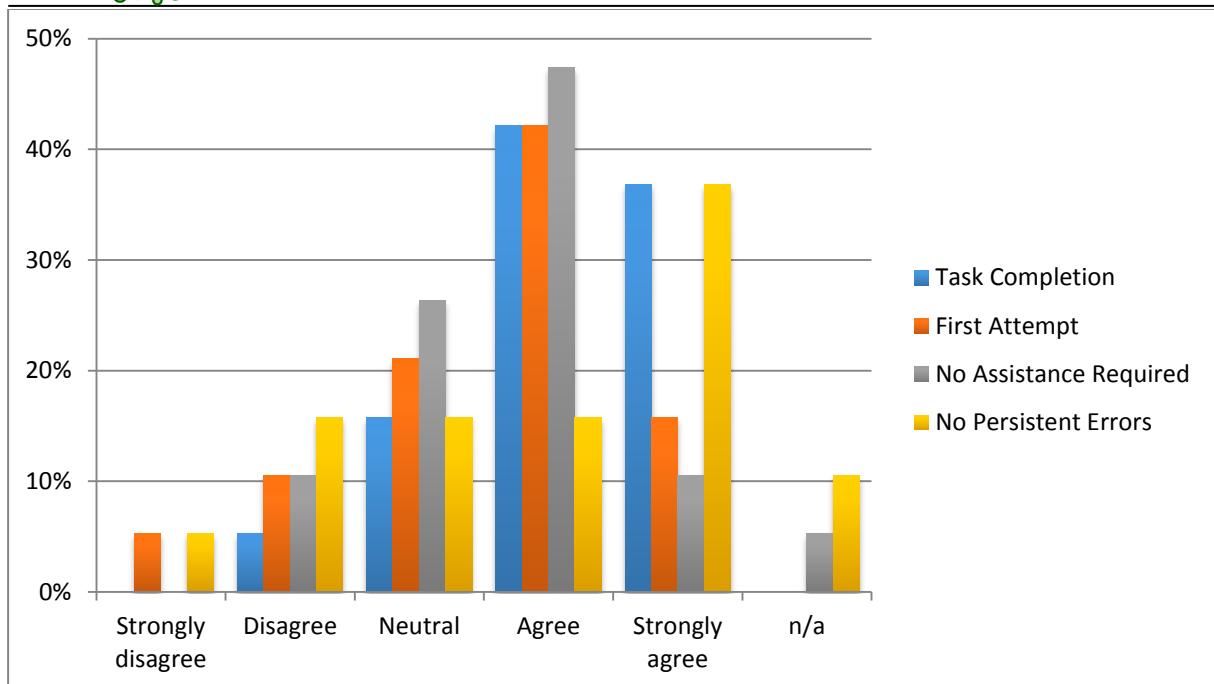


Figure 31: PUC2: General Effectiveness.

b) Efficiency Evaluation

Overall, the system was perceived as easy to understand and easy to use by 69% of evaluation participants agreeing or strongly agreeing on the specific question (Figure 32). When it came to comparing the MULTISENSOR tool with the current method of work or with alternative tools, the majority of users were rather undecided, with 47% of them giving neutral answers or not wanting to answer the question at all.

"I cannot really compare the MULTISENSOR to other ways of performing particular tasks. It is too much of a test setup and the tasks I perform are too much of a daily routine and therefore I perform them too fast to compare the time I needed to the time it took me performing the MULTISENSOR test."

Nevertheless, 42% agreed that the system has timesaving potential compared with the current workflow and 37% thought MULTISENSOR to produce faster results than known alternative tools. Compared with the previous evaluation, the answers for all four questions have slightly improved.

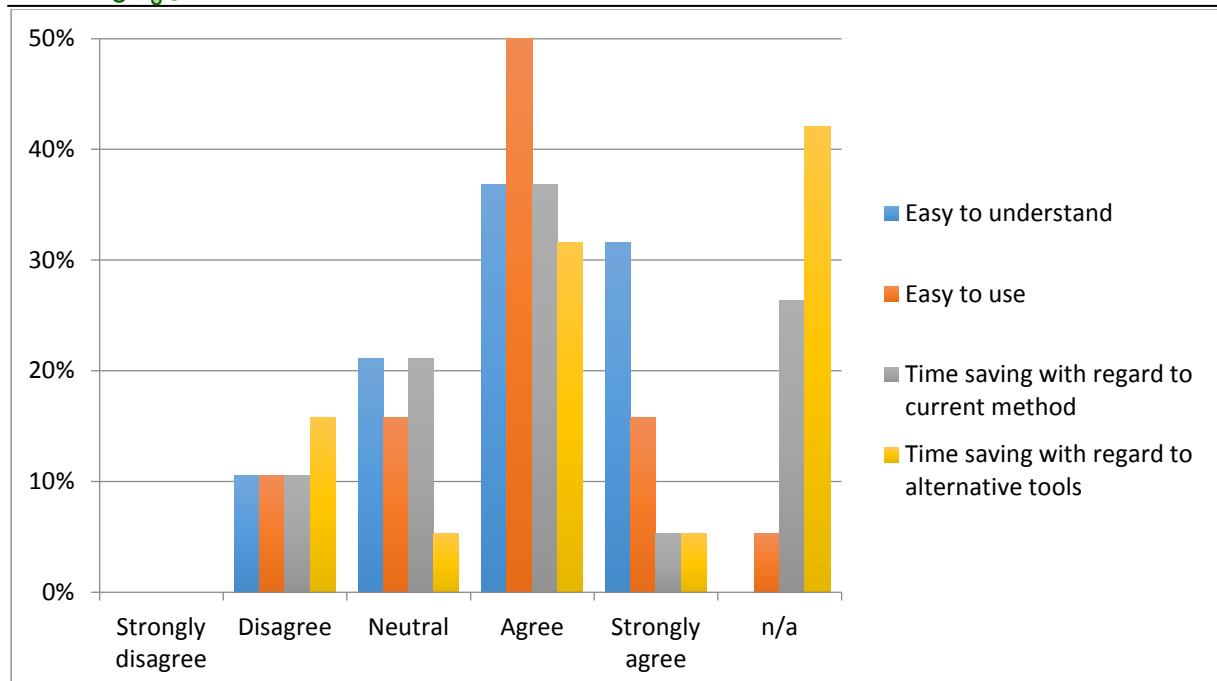


Figure 32: PUC2: Efficiency.

c) Satisfaction Evaluation

The last question of the questionnaire evaluated the overall user satisfaction with the MULTISENSOR PUC2 interface. Again, 63% of participants agreed they felt in control while using the tool and perceived the tool to provide a satisfying experience (Figure 33). 64% were ready to recommend MULTISENSOR to others. The assessment of the tool's usability and its ability to increase productivity produced somewhat lower results. Again, there were some neutral and undecided testers but only a small part of them was outright critical (5 to 10%). Just as was the case with efficiency and effectiveness evaluation, positive answers to all our questions evaluating user satisfaction have increased since the previous evaluation round.

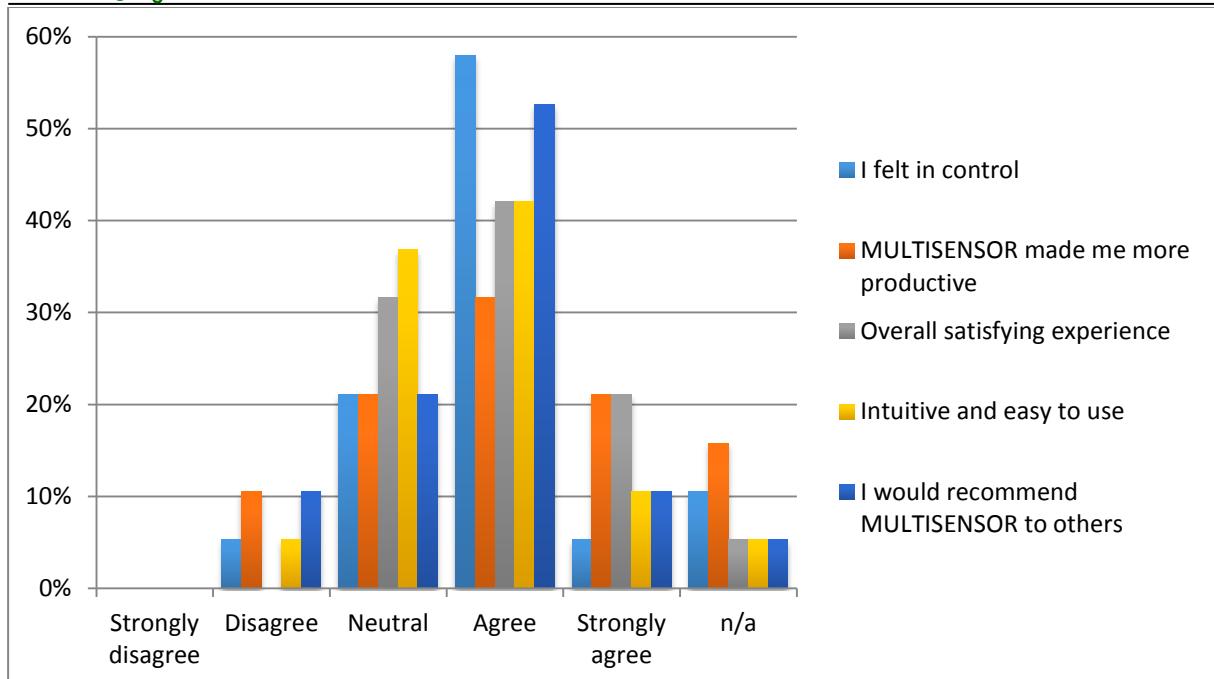


Figure 33: PUC2: Satisfaction.

2.2.5. General Comments

The final item of the evaluation questionnaire asked users to give us their opinion about which functionality of the system they considered as most promising and suitable for further development and subsequent exploitation. The answers to this had no very clear focus as most of the features were mentioned one to three times, the favourite feature being summarisation.

The promising feedback in the evaluation of the 2nd prototype had already prompted pressrelations to integrate both the summarisation and the translation services into their back-office application. These features were integrated with the additional possibility to lengthen or shorten the resulting summary through the use of a slider, and it is possible to obtain an English or German translation instead of the original language summary. Initial feedback from employees is positive especially for the translation. As pressrelations' back-office application allows for an editing of the created summary, remaining quality issues in the resulting summary have less of a weight than in the MULTISENSOR prototype.

2.2.6. Evaluation Analysis and Final Assessment

The third and final evaluation session of the PUC2 MULTISENSOR prototype returned overall improved results with regard to the system's effectiveness, efficiency and user satisfaction, even though the complex media monitoring workflow is only partly covered by the interface. With roughly two thirds of users giving positive feedback for the system, there is however still room for improvement both with regard to features as to usability.

Even though the evaluation was not intended to have a formative approach, our interviewees provided us with additional ideas and requirements during the evaluation

sessions. We also gained good feedback on which of the MULTISENSOR features have a good chance of being exploitable in their current state and which need further development.

Several tools and services, especially summarisation and translation, were taken up very positively, confirming the results of our previous evaluations. Features integrated only more recently such as keyword-based and multi-document summarisation were perceived as potentially useful, but seen to require both further fine-tuning in product development and qualitative improvement.

The evaluation showed that automated text analysis using context information like entities and categories is perceived to have high relevance for the media monitoring market and may provide exploitation potential once an adequate quality of results can be assured.

The features in the influencer section (influence score and community detection) were judged to be highly relevant and helpful to the media monitoring market, being potentially interesting once user requirements to the interface and workflows are adequately met.

Revisiting the user requirements collected for prototype development, most open items and issues had been addressed by the final PUC2 prototype with some minor exceptions that received a lower priority during the development process. Even though the interface does not in its current form have the potential to exhaustively address all workflow requirements in the media monitoring industry, the resulting prototype can now function as a demonstrator for the modules and technological services that can be integrated into media monitoring workflows. The successful integration of the summarisation and translation services into PR's software is a case in point.

2.3. Pilot Use Case 3: SME Internationalisation

2.3.1. Prototype Description and Features

The final prototype development was based on the requirements and its updates from the second evaluation reported in Deliverable 8.4. Mainly, the incorporation of a calculation that suggests countries to the user and also the incorporation of more indicators, which was introduced in the country suggestion section. In addition, summarisation and translation have been made available for Use Case 3. Last, further improvements on the social media visualisation and on GUI development have been carried out.

The initial country indicator module remained unchanged since the UC3 second prototype. The platform shows the indicators divided into four sections: Politics, Economy, Society and Culture (Figure 34).

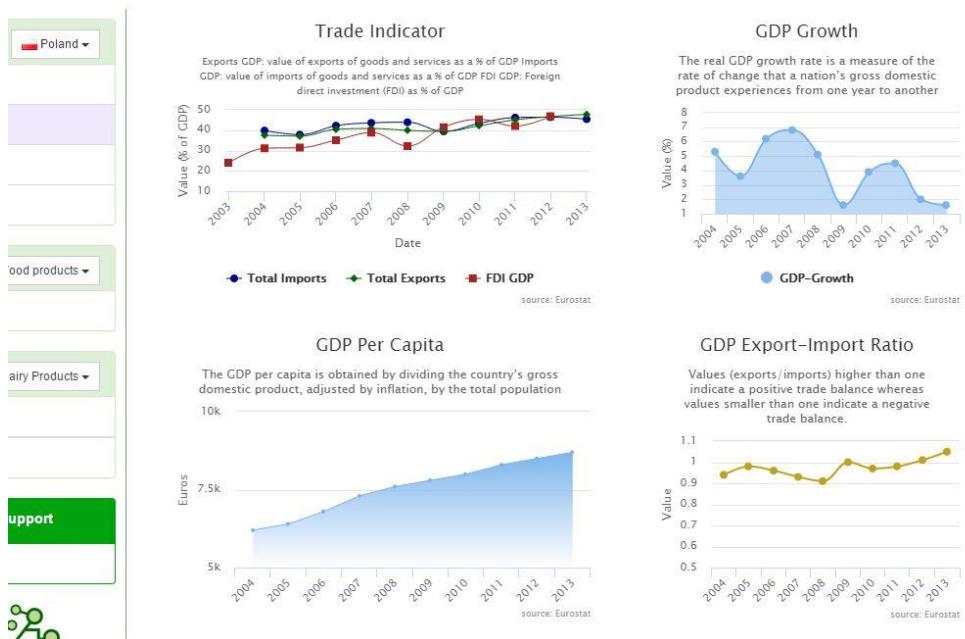


Figure 34: Country Indicators

In the news and articles related to the selected sector and product, summarisation and translation have been incorporated so that the user can relate to the totality of information that is presented to him or her. Furthermore, hybrid search has been included in order to improve the relevance of the articles displayed (Figure 35).

Title	Country	Source	Date
Dairy product prices fall at the latest Global Dairy Trade auction Read the whole summary	IE	agriland.ie	15/03/2016
Summary It also hinted that Whole Milk Powder (WMP) could be up by as much as 7% however, the latest auction showed WMP prices to be down by 0.8%. Rennet Casein posted the lowest decrease (-7.0%) followed by Anhydrous Milk Fat (AMF) which was down by 6.5%. The sixth Global Dairy Trade auction of the year has posted a drop in dairy product prices of 2.9%. This fall comes after dairy produc Read the whole summary			Translate into: de ▾
Specific Concepts: dairy product, Key, Milk Fat, decrease, WMP, grass, cows, last quarter, farmers, production growth, milk production, milk quotas			

Lactose "free" with milk and dairy products? Read the whole summary	IT	noodls.com	11/03/2016
Summary The lactose (lactose) was, at, removed also milk and dairy products is increased under this. All products were indicated and provided with the note as 'lactose free', that the lactose share is less than 0.1 g/100 g. Lactose "free" with milk and dairy products? 25 products were examined, under this different yogurts and fermented milk products, soft, cream cheese and quark, H ful Read the whole summary			Translate into: de ▾
Specific Concepts: buttermilk, quark, samples, shelves, web site, soft cheese, maturation, matter of courses, ingredient, dairy products, lactose intolerance, German population			

Exports of dairy products grow a 50 per cent in the last six years Read the whole summary	ES	Producción Alimentaria	17/03/2016
Summary Organised by DSM, a company of reference in the sector of ingredients for the dairy sector, and Alifarma, the company leader in the distribution of ingredients and food additives, with the participation of Tetra Pak, company in the process and packaging of food, the meeting has had the participation of many of the sector." In this context of potential growth of demand, production Read the whole summary			Translate into: de ▾
Specific Concepts: food additives, company, trends, incorporation, cost of production, technical director, technological innovation, range, dairy products, market share, domestic consumption, tonnes			

Figure 35: Articles displayed for “dairy products” search

The social media crawling has been improved for a better analysis and its visualisation has been also ameliorated. The Twitter users detected are now more relevant for the user search, and the user can click in all of the profiles to directly see their Twitter feed to explore what can be interesting (Figure 36).

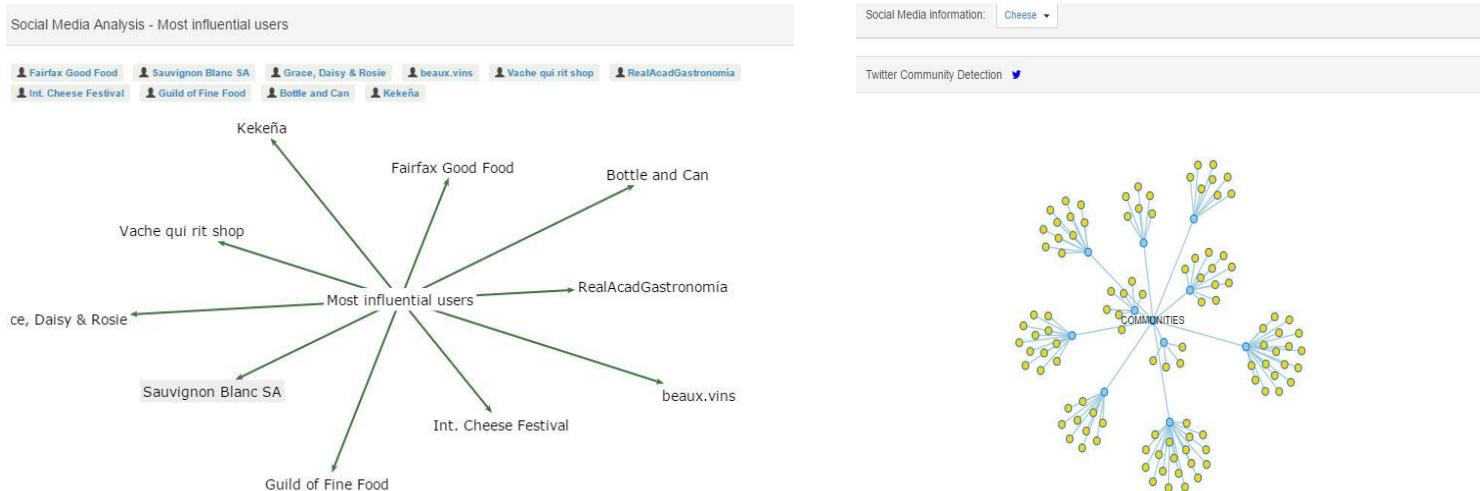


Figure 36: Most influential users and Community Detection display

As stated above, the main novelty of the final prototype is the integration of the Decision Support System. The user needs to select the product, the country of origin and the two countries he/she wants to compare (Figure 37).

Decision Support	
Select your product and country of origin. Then choose two countries to compare and obtain the Decision Support results.	
Step 1: Select product you want to export.	<input type="button" value="Cereals ▾"/>
Step 2: Select your country of origin.	<input type="button" value="Choose a country... ▾"/>
Step 3: Select the two countries you are considering export to.	<input type="button" value="Poland ▾"/> <input type="button" value="Spain ▾"/> <input type="button" value="Submit"/>

Figure 37: Prototype guidance for the Decision Support run

Once the user has selected these features, the system presents the results based on the extended set of indicators (Figure 38), which incorporates specific product export and import data.

Table of indicators			
The selection of the correct country for the international investment depends on a number of indicators. These indicators are very important in order to make a first analysis of the different options that can be presented in a global market.			
#	Indicator	Poland	Germany
6	GDP-Exports of goods and services ⓘ - Economy (%)	47.8 (2013)	50.7 (2013)
7	Balance of trade ⓘ - Economy (ratio)	1.05 (2013)	1.14 (2013)
8	Unemployment ⓘ - Society (%)	10.7 (2014)	5.5 (2014)
9	Total Population ⓘ - Society (people)	38,495,659 (2014)	80,780,000 (2014)
10	Economic Sentiment ⓘ - Society (Index)	99.3 (2014)	103.7 (2014)
11	Internet households ⓘ - Culture (%)	43 (2013)	78 (2013)
12	Ease of doing business ⓘ - Culture (World ranking)	55 (2012)	20 (2012)

Final results			
Final conclusion:	Winner of the comparison:  Germany		
Suggested countries:	Country 1: Netherlands	Country 2: France	Country 3: Italy

Figure 38: Decision Support results display

2.3.2. Evaluation Set-Up

As for the other two Use Cases, we followed a *summative* evaluation focusing on the system usability and the relevance of the modules for the user workflow.

We engaged up to 30 participants, including some members of the User Group, which is a higher number than the previous evaluations.

This time, we also included a question on exploitation to have more elements for its discussion and development.

2.3.3. Task-related evaluation

The evaluation followed the same structure as the Use Case 3 interface so the test persons could go through the platform in a logic manner while responding the questionnaire. Thus, the users were first asked to visualise the different sections of the country indicators and asses. Second, the product and sector information was asked to be analysed both on the news and on the social media analysis. Last, the user had to assess the decision support system. The questionnaire guided the user through the different sections and asked to realise specific actions so that the evaluation process could be smooth.

Generally, the questions for each of the features asked if it the information displayed was easy to understand and if it was relevant for the users work in the SME internationalisation context. Additionally, other questions on how helpful the modules were in relation with comparing countries or analysing markets for internationalisations were included.

For the questions and modules that coincided with the second evaluation round, we will present both results to see the evolution. Concretely, this is the case for the country information and the news and articles questions on understandability and relevance for SME internationalisation.

a) Country information

The country indicators did not experience a significant change in the prototype. Consequently, the results in the final evaluation are very similar from the second evaluation ones. The picture is very positive, as the indicators display is valued well almost by all the participants in the evaluation. Indeed, the neutral answers barely overcome the 10% in some of the cases (Figure 39).

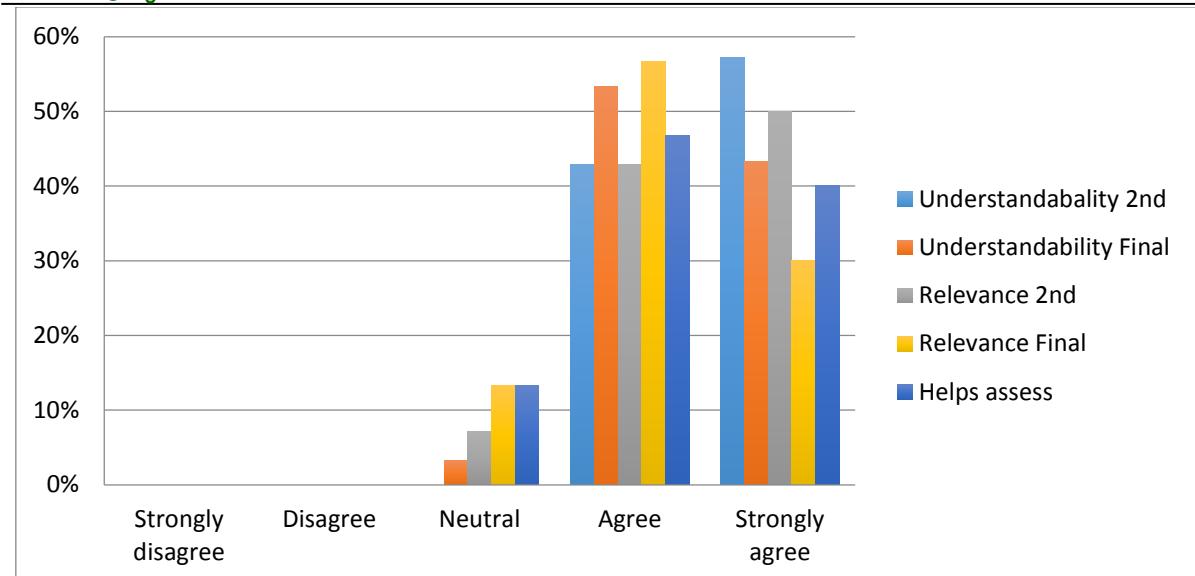


Figure 39: Country information evaluation results

b) News and articles

Regarding the articles display, there was a change in the interface so it looked more similar to the Use Case 1 image. In this sense, we can see an improvement in the understandability of the information displayed. The neutral responses decrease and the strongly agree go up to more than 30% (Figure 40).

Furthermore, there was an improvement in the article crawling queries and hybrid search was incorporated in the article search prioritisation. Here, we observe a significant step forward as the relevance is valued much more positively. In fact, the negative answers practically disappear and the overall testing is much more positive.

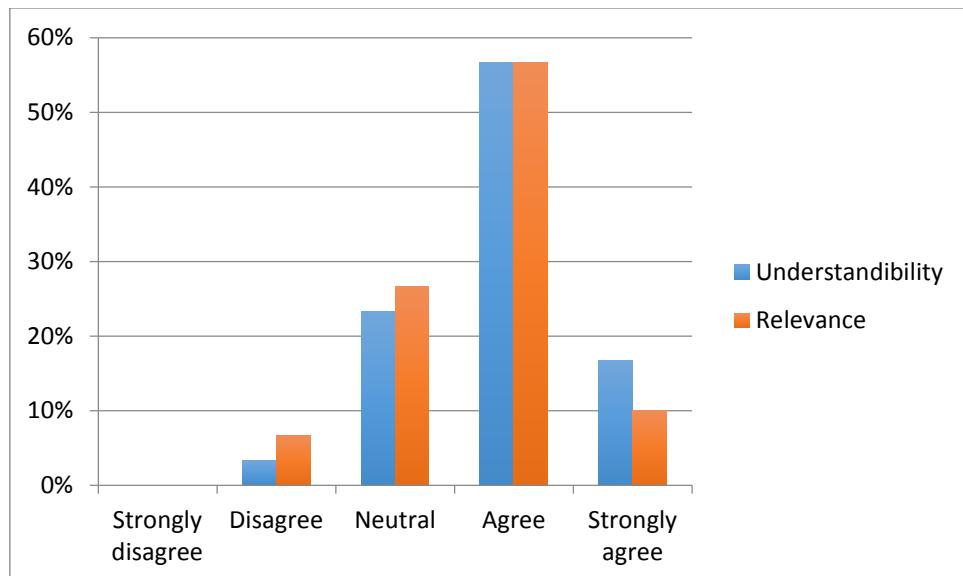


Figure 40: News and articles evaluation results

c) Social Media

The social media analysis that detects Twitter communities and Twitter influencers depending on the selected product was valued quite positively. A vast majority agreed or strongly agreed on the understandability and relevance of the charts displayed (Figure 41). Users valued that they could directly go to the Twitter feed of the selected profiles by clicking on them.

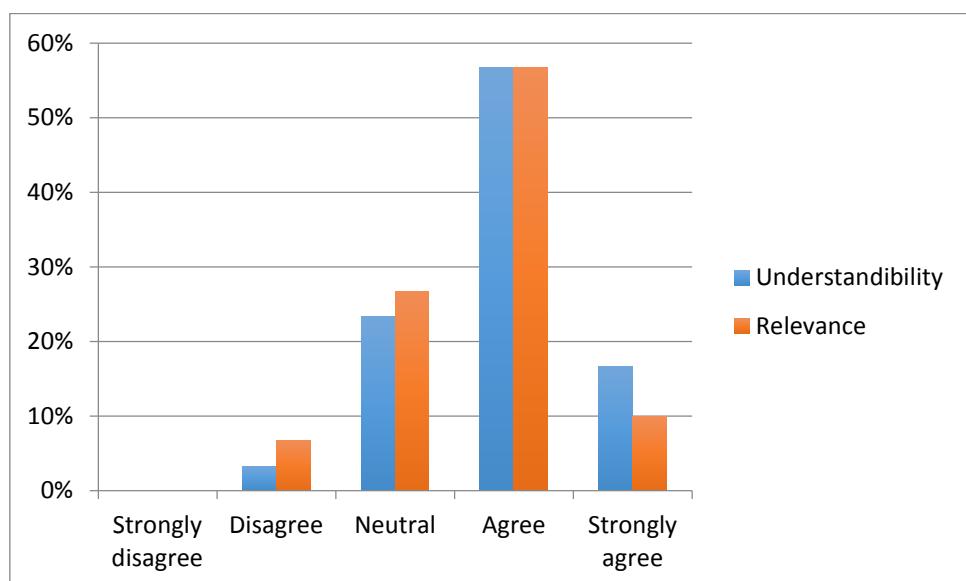


Figure 41: Social Media evaluation results

d) Decision Support

The Decision Support (DS) was valued very positively (Figure 42). Almost all the respondents saw the table of indicators easy to understand, relevant for comparing foreign markets and informing a decision. In addition, the Decision Support results and suggestions also received good feedback on its relation with SME internationalisation.

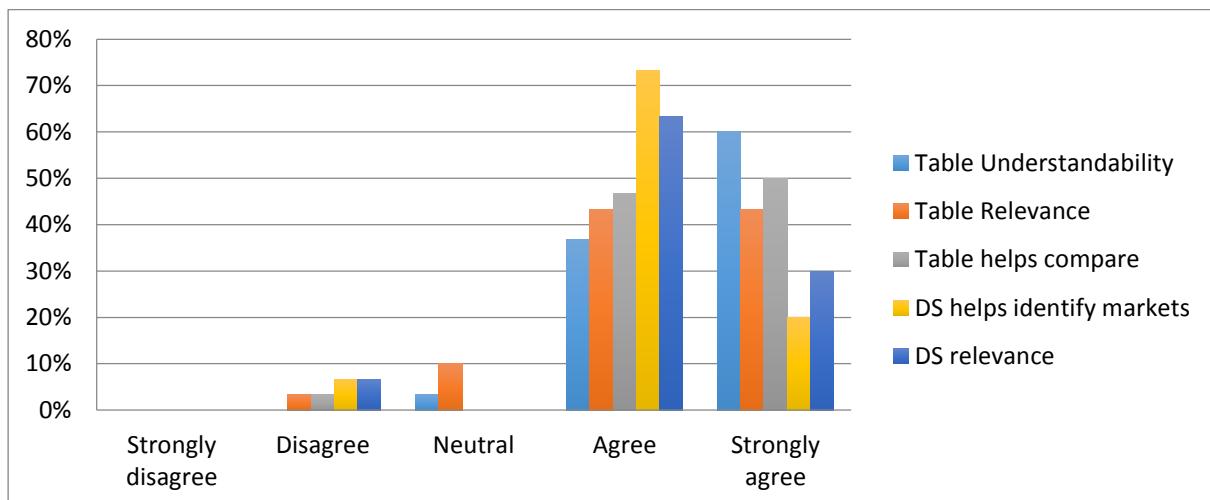


Figure 42: Decision Support evaluation results

2.3.4. Usability testing

a) Effectiveness evaluation

The system did not have persistent errors during the evaluation; thus, the results are very good with regard to effectiveness (Figure 43). The users navigated through the platform with no incidences and completed the tasks with barely any errors.

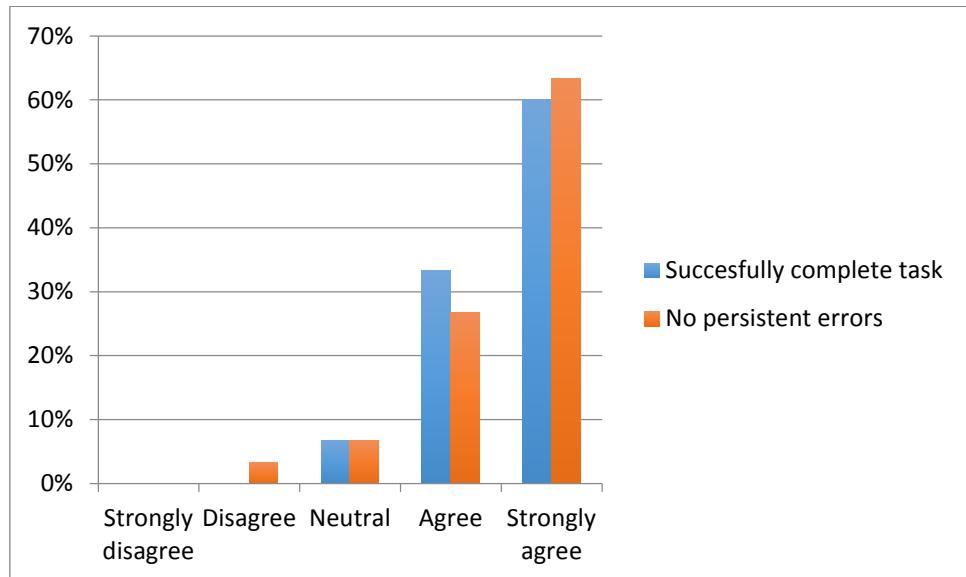


Figure 43: Effectiveness evaluation results

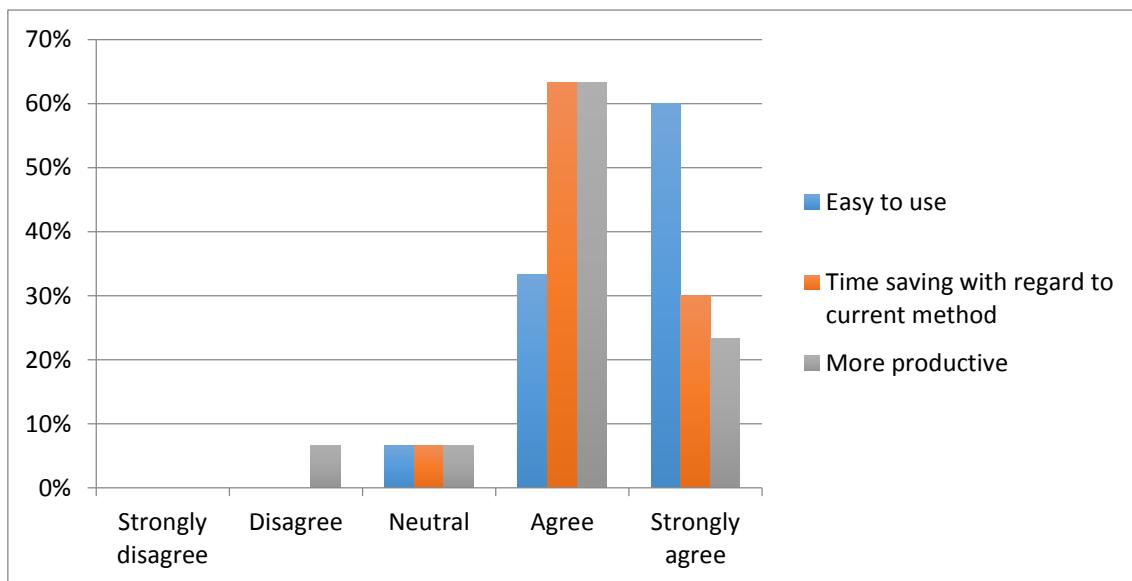


Figure 44: Efficiency evaluation results

b) Efficiency evaluation

Regarding efficiency, the evaluation turned out very positively as well (Figure 44). The users found the platform easy to use and agreed that it would make them more productive when

looking at data from different countries and gathering information to make a decision on which countries have the better conditions for exporting to.

c) Satisfaction evaluation

In terms of satisfaction with the platform GUI, the users were generally happy with the interface and they valued it positively (Figure 45). We see that practically all the users agreed that it was intuitive, that they felt in control and that they found the navigation through the platform a pleasant experience.

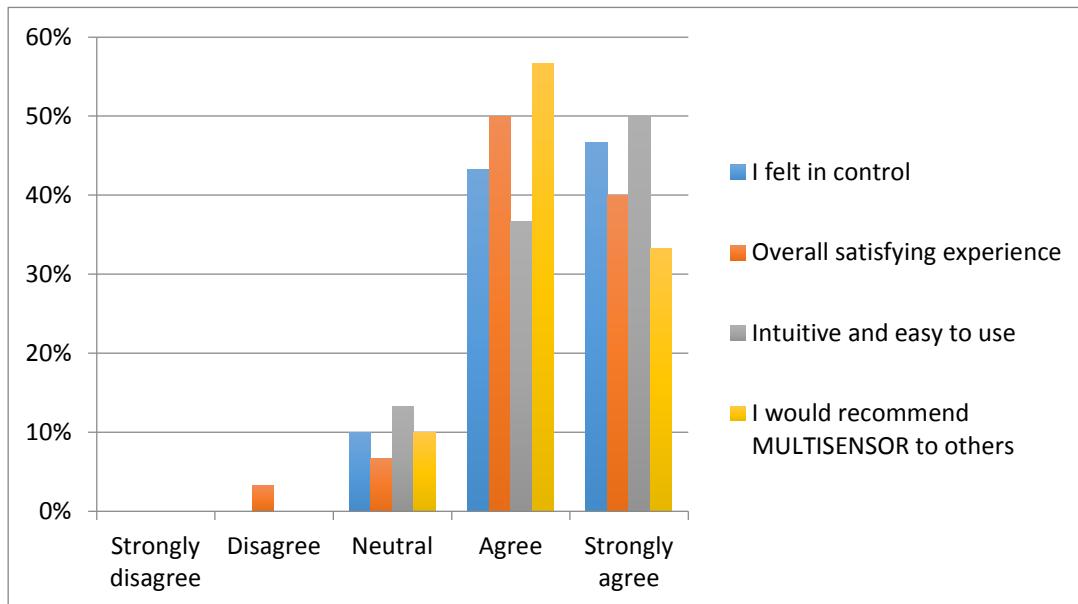


Figure 45: Satisfaction evaluation results

2.3.5. General Comments

The quality feedback from the users suggested some minor modifications or possible additions to the country indicators and a few mentioned that not all the articles were related to market information. All the comments can be consulted in Appendix C.2.

Regarding the answers to the question on exploitability of Use Case 3 features, most of the users found the Decision Support tool as the most promising for commercialisation. Although some mentioned that it needed further development, it was definitely the common growth. A few respondents mentioned the social media analysis as another option as export managers are often not very familiar with them and they could use that help.

2.3.6. Evaluation Analysis and Final Assessment

In all the evaluation results are very positive for the SME internationalisation case. A vast majority of users agreed on the usability and relevance of the features for the Use Case purpose. Particularly, the country indicators and the Decision Support were valued the most. The users found the platform easy to use and with a coherent interface. Remarkably, there was an improvement in the evaluation results of the articles information and display after

the developments carried out during the third year of the project; and, as said, the Decision Support had a good feedback, which was a core element of work of this third year.

In this final evaluation, we got a higher number of participants and we kept involving members of the User Group. In this sense, the Second Open User Day in Barcelona was of good value for Use Case 3 as most of the assistants were export managers with significant experience on SME internationalisation. Thus, there was a valuable exchange between the partners and the attendants to the MULTISENSOR event.

Furthermore, we engaged with some of the participants to further discuss exploitation possibilities for the platform tools and the Decision Support part was the more highlighted, as it is refrained in the questionnaire comments. Nevertheless, a few noted that further development would be needed especially on expanding the products available to have real exploitation prospects. This is reflected and developed in the Exploitation deliverable D9.7.

Overall, the Use Case 3 final prototype interface and modules were well received by the users.

3. FINAL ASSESSMENT OF THE MULTISENSOR SYSTEM AND THE PROJECT RESULTS

3.1. Status Assessment of the Final MULTISENSOR System

In addition to the results from the user-centred evaluation, the MULTISENSOR prototype has been assessed against the expectations formulated in the DoW for Milestone 5. Here, we are referring to the MULTISENSOR system, including the three different use cases.

With regards to the MULTISENSOR system, Milestone 5 marks the successful completion of the third development cycle including an operational architecture that is running under a data security framework. It also includes the final versions of the individual modules:

- Content extraction: advanced multimedia event detection and concept linking.
- User-centric content extraction: sentiment, context extraction and social media mining using advanced techniques.
- Content integration and retrieval: semantic integration, topic detection and tracking, advanced multimodal indexing and retrieval
- Reasoner and decision support: Final version of the decision support system integrating the advanced reasoning techniques.
- Information production: integrated summarisation system.

3.2. Assessment of the Project Results

The project achieved good results in the final evaluation of the three different prototypes. Generally, all the modules available were rated positively by the users that participated in the evaluation. Particularly, automatic summarisation and the Decision Support system received promising results and have the better possibilities for exploitation. In fact, pressrelations integrated both MULTISENSOR's summarisation and translation into its software. Furthermore, the system as a whole, which was also the focus in this last summative evaluation, showed good acceptance among the evaluators.

In relation with WP8 Objectives defined in the deliverable D1.1 and reported in D1.2, the project has met with the initial expectations. More concretely, the project has achieved its highest expectation for objective 1 by creating several solutions that considerably facilitate the work of professionals of the three use case areas: journalism, commercial media monitoring and SME internationalisation. Similarly, MULTISENSOR also met the highest expectation for objective 2, as the assessment on the system's usefulness and usability was very positive. The users, by a large majority, saw the online tools as easy to use and navigate through for the specific tasks of the three different professional workflows.

Objective 3 has been achieved with the highest expectations, as no remedial actions have been necessary for any of the project's WPs. In regard with Objective 4, the metrics for all objectives were available at least 14 days ahead of critical points; thus, the lowest expectation has been met. Last, the project achieved the lowest expectation for Objective 5 as MULTISENSOR developed modules that improved existing worflows and performed better

than other modules available. In this sense, pressrelation incorporating MULTISENSOR's software into its workflow is a good example, as well as PIMEC's initial use of the Decision Support System. Nevertheless, for an integrated system with a unique selling proposition, further development would be needed. Table 4, summarizes the expected and achieved objectives for WP8 based on the indicators defined in D1.1 and D1.2.

Globally, the initial requirements of the different use cases and the developments and adaptations that have been made throughout the project have been met. In this sense, the three platforms cover the user requirements and are useful for the professional tasks that were meant to. Further development would enrich the platforms and give them a more robust functionality that would ease the exploitation plans.

#	Highest Expectation	Achievement
1	The MULTISENSOR project creates one or several solutions and modules that considerably facilitate the work of journalists, commercial media monitors and business managers with regard to their specific working environment.	Achieved. MULTISENSOR created several solutions that considerably facilitate the work of professionals of the three use case areas: journalism, commercial media monitoring and SME internationalisation
2	Very good assessment (in line with the Likert scale) of one or more test users in terms of usefulness, and usability for each of the identified user types.	Achieved. The majority of users saw the online tools as easy to use and navigate through for the specific tasks of the three different professional workflows.
3	No remedial actions are necessary for any of the WPs.	Achieved. No remedial actions have been necessary for any of the project's WPs
4	The metrics are available at least 1 month ahead of critical time points such as the Milestones at which the corresponding techniques have to be operational.	Partially achieved. The metrics for all objectives were available at least 14 days ahead of critical points
5	The MULTISENSOR project creates an integrated system with a unique selling proposition compared to existing workflows and systems or tools that are already available in the market. 100% fulfillment of the test protocol (D8.1).	Partially achieved. MULTISENSOR developed modules that improved existing workflows and performed better than other modules available. Nevertheless, for an integrated system with a unique selling proposition, further development would be needed.

Table 4: Expected and Achieved Objectives for WP8

4. CONCLUSION

The final evaluation showed good results for the three MULTISENSOR use cases developed. For this final evaluation we followed a summative approach, as opposed to a formative one, as it was the last evaluation of the project. Hence, we needed an evaluation of the usability of the platform and its modules for the professional use cases rather than an evaluation that produced hints and directions for future developments. Overall, we involved a higher number of participants in this third and final evaluation round than in previous ones and we continued the involvement of members of the User Group.

For the journalistic case, the automatic summarisation module received remarkably positive feedback. In addition, translation had also good results, improving significantly from previous evaluations. In general, the platform as a whole was valued as useful for a journalistic professional use by a vast majority of the users. In this sense, the analysis that the platform runs on the selected articles was positively assessed. Nevertheless, based on the overall evaluators' feedback and the nature of the evaluation, we acknowledge that some development would be needed to make a real case for exploitability. Having said that, summarisation was highly ranked and seen as very promising in terms of exploitation possibilities; this confirms the project initial hypothesis on the prospects of use case 1 modules.

Regarding the commercial media monitoring case, the evaluation focused on discovering how the MULTISENSOR platform would be accepted at different points of a simulated near-real-life workflow. Generally, the evaluation turned out significantly positive too, particularly for summarisation, translation and the analysis section. For this use case, there was an improvement on the system's effectiveness and efficiency results. In addition, automated text analysis was valued as very relevant for media monitoring. In all, the current platform does not cover all the media monitoring steps workflow but it does have specific modules that can successfully be integrated for this purpose. Moreover, pressrelations has integrated MULTISENSOR translation and summarisation tools into its own software.

For the SME internationalisation case, the results were also very positive. The different modules were seen as relevant for the professional task. In addition, there was an improvement in the news content and visualisation as a result of the developments carried out in the third year of the project. In this sense, the Decision Support feature, which was finalised during this third year, received very positive feedback and was signalled as the main element for a future exploitation.

Generally, the GUI and the interface of the different modules were seen as user-friendly and easy to use. Despite having quite differentiated platforms, the users from the different professional areas were satisfied with the MULTISENSOR usability look and feel. The users did not experience any persistent errors and were able to finish their task-oriented evaluations without any major inconveniences.

5. REFERENCES

Barnum, Carol M. 2011. Usability Testing Essentials

ISO 9241-11. 1988. Ergonomic requirements for office work with visual display terminals. Part 11 Guidance on usability.

APPENDICES

A. Pilot Use Case 1: Journalism

A.1. Questionnaire

MULTISENSOR Final Evaluation

General Information

1. Please select your age.

- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55+

2. Please indicate your gender

- Female
- Male

3. Please state the country you live in

4. Please tell us your native language(s)

5. Please state your current occupation

- Journalist
- Researcher
- Other:

6. If you are a journalist, please specify which category best fits your main role

- Editor/Agenda setting - choosing and selecting stories
- Writer/Subeditor - writing or assembling stories from multiple sources
- Researcher - specialised in researching topics and background information
- Social Media Specialist/Digital Analyst
- Other:

Professional User Test

First steps

- Before starting, you need to log in (for the purpose of this evaluation only). In order to do so, please click "GO" on the landing page.
- On the next page you will find a log-in icon in the top right-hand corner.
- Please use the credentials that you have received in order to log in.
- After you have done this, please click on the "home" - button in order to return to the landing page.
- Now, please read the scenario, perform the following tasks and evaluate the results.

The Scenario: Imagine you're a journalist assigned with the task to find out more about **Energy policy in the UK and the US**.

In order to do so, you want to find relevant articles from multiple sources. You're on a tight schedule and have to include articles from different countries.

Your editorial department uses the Multisensor platform, a tool to help you make quicker decisions on what articles could be useful for your work.

Your goal is to **create a personal portfolio of at least 5 relevant articles**.

The Tasks

Task 1: Start your query

You are looking for five relevant articles including the keywords “energy policy” in English, that have been published between January 1st, 2016 and today. Please perform this query in the application.

Query page

7. The automatically suggested search terms below the search box that appear when typing the query are useful for fulfilling the task

1 2 3 4 5

Strongly disagree Strongly agree

Results page

8. The listings of people, places and organisations in the results are useful for fulfilling the task.

1 2 3 4 5

Strongly disagree Strongly agree

Task 2: Create a personal dossier

Have a look at the list of results from your query. Please identify relevant articles by using the following features that the MULTISENSOR system provides:

a) Summarisation

b) Translation

c) In-depth analysis

Please use all of these features when analysing articles as you will be asked to assess their performance later. If an article appears to be relevant, add the article to your dossier. Continue this process until your dossier contains a minimum of five different articles.

Now, please answer the following questions:

a) Summarisation

9. The summaries are useful for quickly deciding on the relevance of articles.

1 2 3 4 5

Strongly disagree Strongly agree

10. The quality of the summaries is adequate for fulfilling the task.

1 2 3 4 5

Strongly disagree Strongly agree

11. The list of concepts from the individual article is useful for my assessment of this article.

1 2 3 4 5

Strongly disagree Strongly agree

b) Translation

Please be aware that this service is not aiming for a perfect translation, but is supposed to support you in assessing whether a foreign language article might be relevant.

12. The translation is useful for quickly deciding on the relevance of an article.

1 2 3 4 5

Strongly disagree Strongly agree

c) In-depth Analysis

13. The list of entities found in the text is useful for quickly deciding on the relevance of an article.

1 2 3 4 5

Strongly disagree Strongly agree

14. The wordcloud with key concepts from the text is useful for quickly deciding on the relevance of an article.

1 2 3 4 5

Strongly disagree Strongly agree

15. The list of related articles is relevant for expanding the research and identifying new articles.

1 2 3 4 5

Strongly disagree Strongly agree

Further remarks?

16. Please write down any further remarks and suggestions you might have.

Task 3: Use the multimedia search

There is an additional feature for articles including multimedia items in the advanced search. Please go back to the list of results and repeat your search for articles including multimedia items.

17. The provided multimedia information is useful for further research.

1	2	3	4	5
Strongly disagree <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Strongly agree				

Task 4: The portfolio analysis

Now open your portfolio and run the portfolio analysis. Please assess the results by answering the following questions:

18. The list of entities is useful for assessing the relevance of the chosen articles.

1	2	3	4	5
Strongly disagree <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Strongly agree				

19. The keyword cloud is useful for assessing the relevance of the chosen articles.

1	2	3	4	5
Strongly disagree <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Strongly agree				

20. The list of related articles, based on related topics, is relevant for expanding the research and identifying new articles.

1	2	3	4	5
Strongly disagree <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> Strongly agree				

Further remarks?

21. Please write down any further remarks and suggestions you might have

Overall assessment and feedback

Effectiveness



22. I was able to successfully complete the scenario.

1 2 3 4 5

Strongly disagree Strongly agree

23. I could complete the scenario on the first attempt.

1 2 3 4 5

Strongly disagree Strongly agree

24. I could complete the scenario without external assistance.

1 2 3 4 5

Strongly disagree Strongly agree

25. I did not notice any persistent errors while using the application.

1 2 3 4 5

Strongly disagree Strongly agree

Efficiency

26. It did not take me a lot of time to understand MULTISENSOR and learn about its functionalities.

1 2 3 4 5

Strongly disagree Strongly agree

27. The MULTISENSOR system was easy to use and the main functionalities were easy to find.

1 2 3 4 5

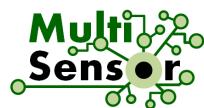
Strongly disagree Strongly agree

28. It did not take me a lot of time to perform the scenario compared to my current method of handling similar tasks.

1 2 3 4 5

Strongly disagree Strongly agree

29. It did not take a lot of time to perform the scenario compared to the use of alternative tools that I have already used.



1 2 3 4 5

Strongly disagree Strongly agree

Satisfaction

30. I felt in control when I used MULTISENSOR for finding relevant articles.

1 2 3 4 5

Strongly disagree Strongly agree

31. It made me more productive when I used MULTISENSOR for finding relevant articles.

1 2 3 4 5

Strongly disagree Strongly agree

32. The use of MULTISENSOR was overall a satisfying experience.

1 2 3 4 5

Strongly disagree Strongly agree

33. The MULTISENSOR interface is intuitive and easy to use.

1 2 3 4 5

Strongly disagree Strongly agree

34. I would recommend MULTISENSOR to a colleague or a friend.

1 2 3 4 5

Strongly disagree Strongly agree

Wrap-up

35. Which functionality/functionalities of the MULTISENSOR system do you consider as most promising and suitable for further development and subsequent exploitation?

Please write down your ideas.

36. Further remarks?

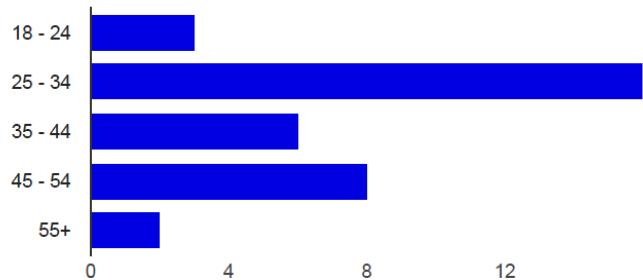
Please write down any further remarks and suggestions you might have.

A.2. Evaluation Summary

Summary

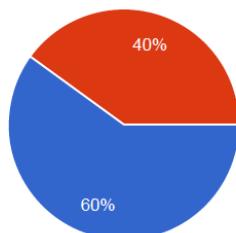
General Information

Please select your age



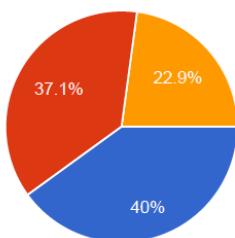
18 - 24	3	8.6%
25 - 34	16	45.7%
35 - 44	6	17.1%
45 - 54	8	22.9%
55+	2	5.7%

Please indicate your gender



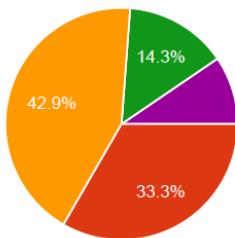
Female	21	60%
Male	14	40%
Female	21	60%
Male	14	40%

Please state your current occupation



Journalist	14	40%
Researcher	13	37.1%
Other	8	22.9%
Journalist	14	40%
Researcher	13	37.1%
Other	8	22.9%

If you are a journalist, please specify which category best fits your main role

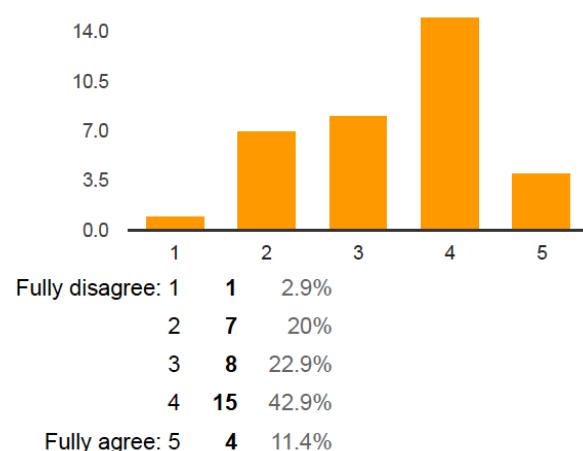


Editor/Agenda setting - choosing and selecting stories	0	0%
Writer/Subeditor - writing or assembling stories from multiple sources	7	33.3%
Researcher - specialised in researching topics and background information	9	42.9%
Social Media Specialist/Digital Analyst	3	14.3%
Other	2	9.5%

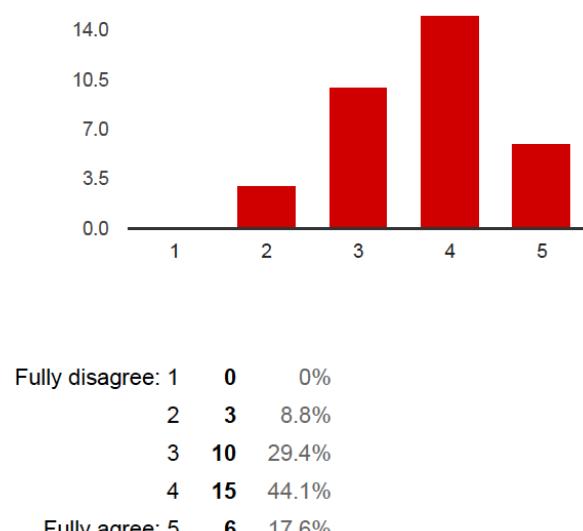
The Tasks

Task 1: Start your query

Query page



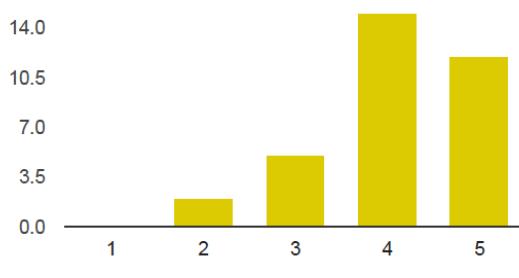
Results page



Task 2: Create a personal dossier

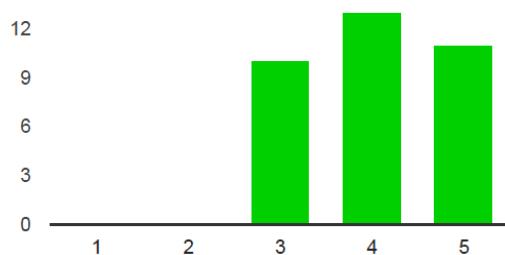
a) Summarisation

The summaries are useful for quickly deciding on the relevance of articles.



Fully agree: 5 12 35.3%
Fully disagree: 1 0 0%
2 2 5.9%
3 6 14.7%
4 15 44.1%
Fully agree: 5 12 35.3%

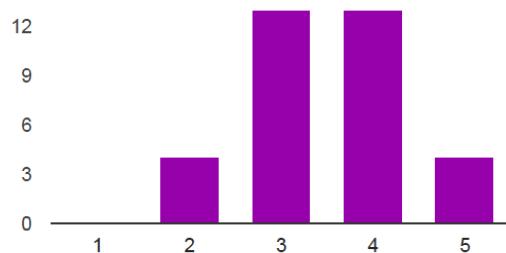
The quality of the summaries is adequate for fulfilling the task.



Fully disagree: 1 0 0%
2 0 0%
3 10 29.4%
4 13 38.2%
Fully agree: 5 11 32.4%

The list of concepts from the individual article is useful for my assessment of this article.

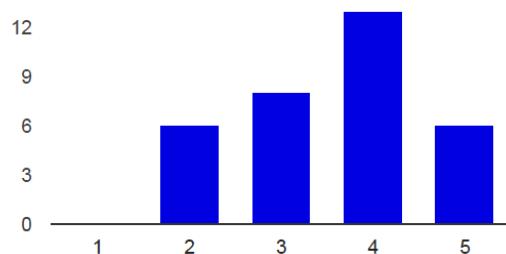
The list of concepts from the individual article is useful for my assessment of this article.



Fully disagree: 1 0 0%
 2 4 11.8%
 3 13 38.2%
 4 13 38.2%
Fully agree: 5 4 11.8%

b) Translation

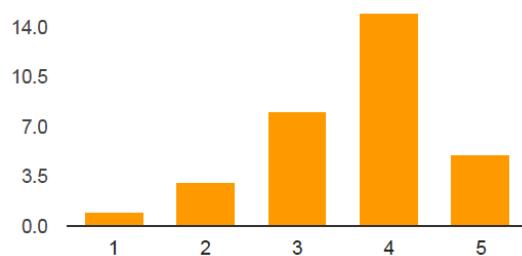
The translation is useful for quickly deciding on the relevance of an article.



Fully disagree: 1 0 0%
 2 6 18.2%
 3 8 24.2%
 4 13 39.4%
Fully agree: 5 6 18.2%

c) In-depth Analysis

The list of entities found in the text is useful for quickly deciding on the relevance of an article.



Fully disagree: 1 1 3.1%

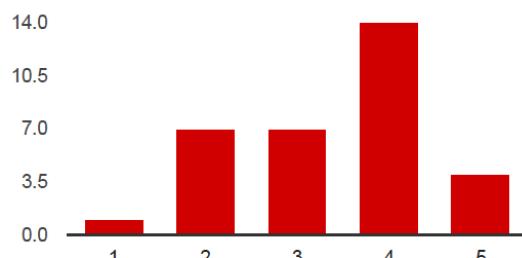
 2 3 9.4%

 3 8 25%

 4 15 46.9%

Fully agree: 5 5 15.6%

The wordcloud with key concepts from the text is useful for quickly deciding on the relevance of an article.



Fully disagree: 1 1 3%

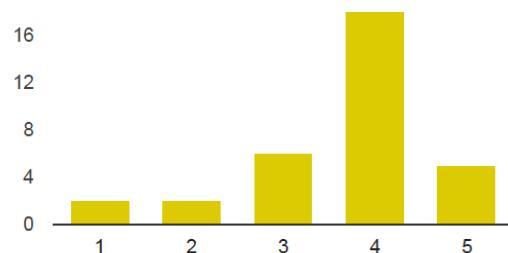
 2 7 21.2%

 3 7 21.2%

 4 14 42.4%

Fully agree: 5 4 12.1%

The list of related articles is relevant for expanding the research and identifying new articles.



Fully disagree: 1 **2** 6.1%
 2 **2** 6.1%
 3 **6** 18.2%
 4 **18** 54.5%
Fully agree: 5 **5** 15.2%

Further remarks?

Highlight main experts or protagonists or organizations mentioned

initial summary, even of a German text is useful. translation of that summary back into German is hard to read.

The interface is a bit too crowded. I would suggest a clearer hierarchy between the most important elements and the supporting elements. Right now the interface is too busy.

Summaries are really helpful; some of the related articles were not related

I could not find any article about US energy policy.

Translation: a button to confirm the language would be useful

Στο cloud να κάνω κλικ στις λέξεις

The word cloud contains a lot of non content words

Entities should add surnames. Some are non relevant. Key point: updating info regularly

Cloud provides common nonsense words which give no clues.

Highlight main experts or protagonists or organizations mentioned

initial summary, even of a German text is useful. translation of that summary back into German is hard to read.

The interface is a bit too crowded. I would suggest a clearer hierarchy between the most important elements and the supporting elements. Right now the interface is too busy.

Summaries are really helpful; some of the related articles were not related

I could not find any article about US energy policy.

Translation: a button to confirm the language would be useful

Στο cloud να κάνω κλικ στις λέξεις

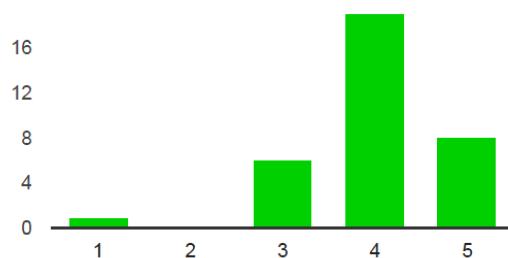
The word cloud contains a lot of non content words

Entities should add surnames. Some are non relevant. Key point: updating info regularly

Cloud provides common nonsense words which give no clues.

Task 3: Use the multimedia search

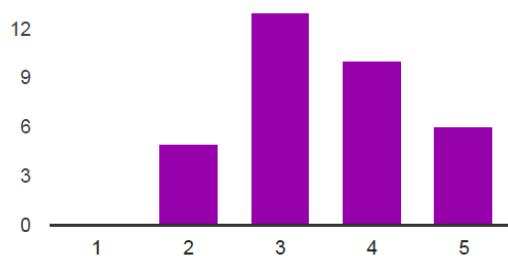
The provided multimedia information is useful for further research.



Fully disagree: 1 1 2.9%
 2 0 0%
 3 6 17.6%
 4 19 55.9%
Fully agree: 5 8 23.5%

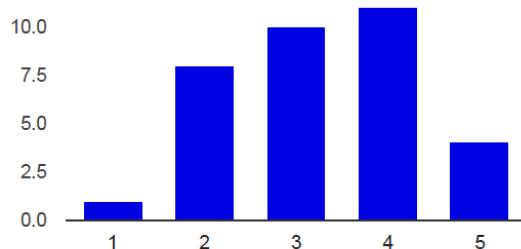
Task 4: The portfolio analysis

The list of entities is useful for assessing the relevance of the chosen articles.



Fully disagree: 1 0 0%
 2 5 14.7%
 3 13 38.2%
 4 10 29.4%
Fully agree: 5 6 17.6%

The keyword cloud is useful for assessing the relevance of the chosen articles.



Fully disagree: 1 **1** 2.9%

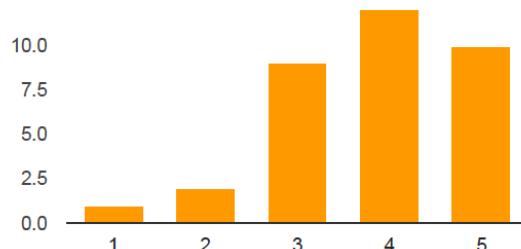
2 **8** 23.5%

3 **10** 29.4%

4 **11** 32.4%

Fully agree: 5 **4** 11.8%

The list of related articles, based on related topics, is relevant for expanding the research and identifying new articles.



Fully disagree: 1 **1** 2.9%

2 **2** 5.9%

3 **9** 26.5%

4 **12** 35.3%

Fully agree: 5 **10** 29.4%

Further remarks?

Showing which multimedia item is part of the article. Keyword cloud is a good idea but prefer only nouns. Related articles in the "run analysis" should be focus on the topic and geolocation. Entity with full name

Helps to find common elements. Multimedia search confusing. It's not clear whether this option is actually enabled, since the multimedia items only appear in the detailed

summary

I am missing a clear "work direction". The results in the portfolio are quite good, but how can I now accomplish something towards my reporting goals?

First I didn't know what extras the multimedia would bring. Couldn't find them intuitively.

The Portfolio Icon was not easy to be found.

Δεν μπορώ να καταλάβω γιατί υπάρχει σχέση στα σχετικά άρθρα αλλά και με ποιο τρόπο
Summarisation is really helpful. I have only checked the translations into German but they were quite inconsistent in quality.

I find the Entities information redundant and less reliable than Persons, Locations and Organisation

I need more context on entities and word cloud

Showing which multimedia item is part of the article. Keyword cloud is a good idea but prefer only nouns. Related articles in the "run analysis" should be focus on the topic and geolocation. Entity with full name

Helps to find common elements. Multimedia search confusing. It's not clear whether this option is actually enabled, since the multimedia items only appear in the detailed summary

I am missing a clear "work direction". The results in the portfolio are quite good, but how can I now accomplish something towards my reporting goals?

First I didn't know what extras the multimedia would bring. Couldn't find them intuitively.

The Portfolio Icon was not easy to be found.

Δεν μπορώ να καταλάβω γιατί υπάρχει σχέση στα σχετικά άρθρα αλλά και με ποιο τρόπο
Summarisation is really helpful. I have only checked the translations into German but they were quite inconsistent in quality.

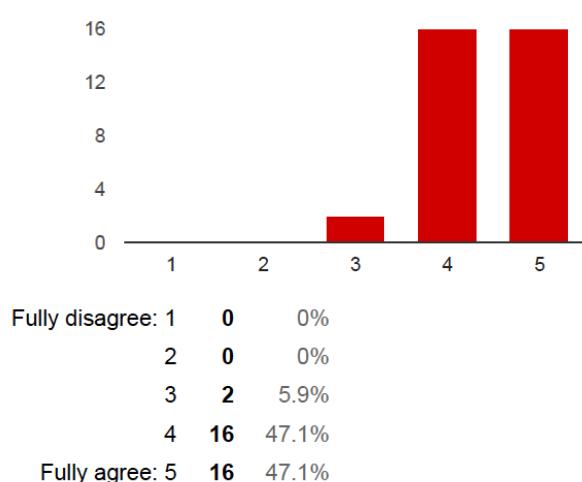
I find the Entities information redundant and less reliable than Persons, Locations and Organisation

I need more context on entities and word cloud

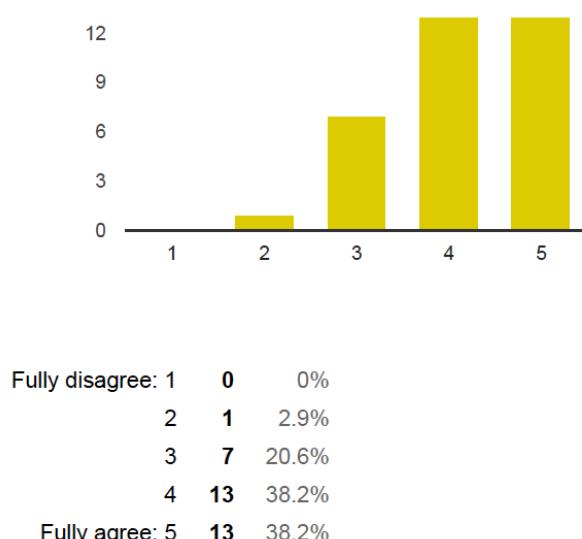
Overall assessment and feedback

Effectiveness

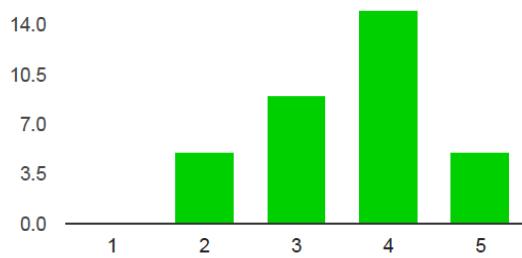
I was able to successfully complete the scenario.



I could complete the scenario on the first attempt.

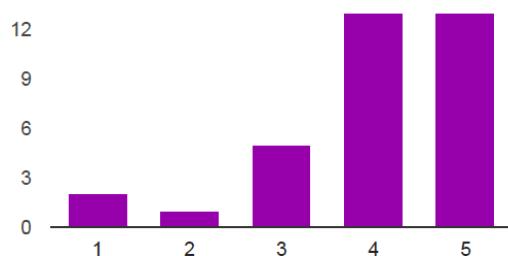


I could complete the scenario without external assistance.



Fully disagree: 1 0 0%
 2 5 14.7%
 3 9 26.5%
 4 15 44.1%
Fully agree: 5 5 14.7%

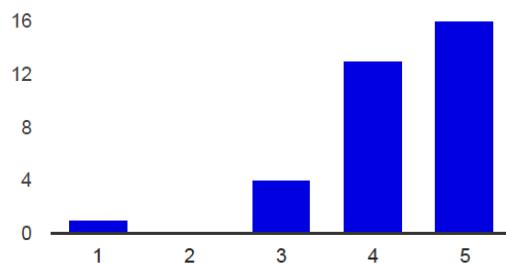
I did not notice any persistent errors while using the application.



Fully disagree: 1 2 5.9%
 2 1 2.9%
 3 5 14.7%
 4 13 38.2%
Fully agree: 5 13 38.2%

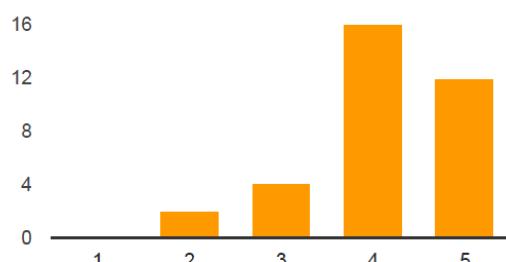
Efficiency

It did not take me a lot of time to understand MULTISENSOR and learn about its functionalities.



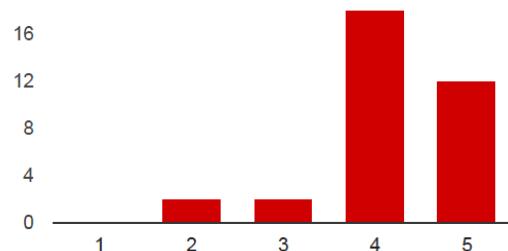
Fully disagree: 1 **1** 2.9%
 2 **0** 0%
 3 **4** 11.8%
 4 **13** 38.2%
Fully agree: 5 **16** 47.1%

The MULTISENSOR system was easy to use and the main functionalities were easy to find.



Fully disagree: 1 **0** 0%
 2 **2** 5.9%
 3 **4** 11.8%
 4 **16** 47.1%
Fully agree: 5 **12** 35.3%

It did not take me a lot of time to perform the scenario compared to my current method of handling similar tasks.



Fully disagree: 1 0 0%

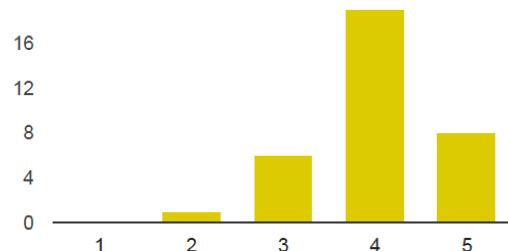
2 2 5.9%

3 2 5.9%

4 18 52.9%

Fully agree: 5 12 35.3%

It did not take a lot of time to perform the scenario compared to the use of alternative tools that I have already used.



Fully disagree: 1 0 0%

2 1 2.9%

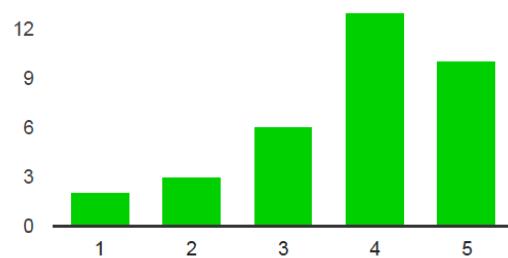
3 6 17.6%

4 19 55.9%

Fully agree: 5 8 23.5%

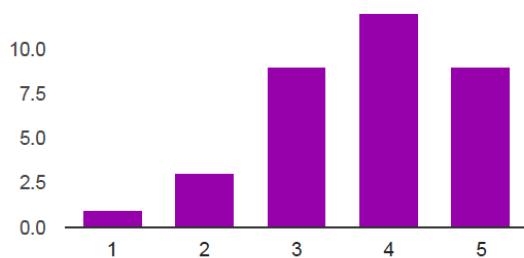
Satisfaction

I felt in control when I used MULTISENSOR for finding relevant articles.



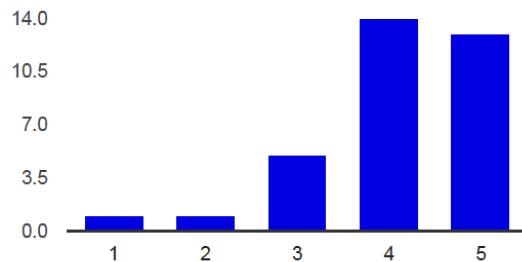
Fully disagree: 1 2 5.9%
 2 3 8.8%
 3 6 17.6%
 4 13 38.2%
Fully agree: 5 10 29.4%

It made me more productive when I used MULTISENSOR for finding relevant articles.



Fully disagree: 1 1 2.9%
 2 3 8.8%
 3 9 26.5%
 4 12 35.3%
Fully agree: 5 9 26.5%

The use of MULTISENSOR was overall a satisfying experience.



Fully disagree: 1 1 2.9%

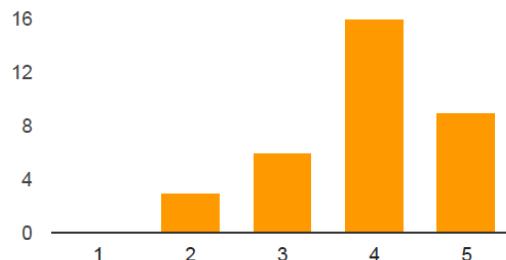
2 1 2.9%

3 5 14.7%

4 14 41.2%

Fully agree: 5 13 38.2%

The MULTISENSOR interface is intuitive and easy to use.



Fully disagree: 1 0 0%

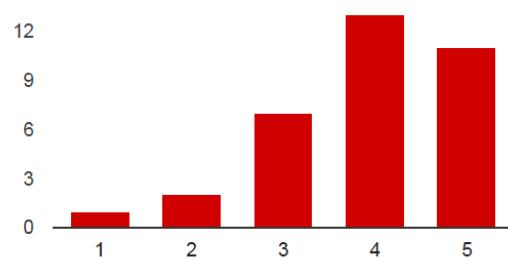
2 3 8.8%

3 6 17.6%

4 16 47.1%

Fully agree: 5 9 26.5%

I would recommend MULTISENSOR to a colleague or a friend.



Fully disagree: 1 **1** 2.9%
 2 **2** 5.9%
 3 **7** 20.6%
 4 **13** 38.2%
Fully agree: 5 **11** 32.4%

Wrap-up

Which functionality/functionalities of the MULTISENSOR system do you consider as most promising and suitable for further development and subsequent exploitation?

Related articles are helping to find more specific content. But next to the "Back"-Button should be an option to go to the query results again. Helpful to look for articles in native language. Summary and translation is saving time of selecting.

The summarization and the extraction of common elements in the dossier.

General idea of finding, saving content, etc is strong

The summarisation tool is very promising; the translation tool is helpful but quality could be better.

Summarization of lengthy text Translations if they were of better quality

the summarization tool was pretty good. it provided a uniform glimpse on what was to be expected from an article. the related content was shown in German language. I liked it, but I was a bit puzzled about the language.

Relates Articles,

Summaries and summary translations could be hugely useful if they were much better.

Finding relevant asticles in many languages

Perhaps the provision of more in detail insights related to the media monitoring strand (and more precisely regarding influencers' data)

The summarizing of articles.

the sentiment / polarity analysis, while also the transcription service was very interesting

Τα θετικά και αρνητικά κείμενα για ένα θέμα σε σχέση με την κλίμακα.

transcription, sentimentality

better link between the different articles with similar theme

related media, transcriprion and run in depth semantic analysis

In depth semantic analysis

The summarisation tool is very helpful.

The translation in other languages

Portfolio analysis

Text summarization, translation and multimedia analysis

Summarisation

Topic and event detection is very interesting

Potfolio analysis

There are external links to find more about entities and ways to use word cloud and entities to filter articles.

Related articles are helping to find more specific content. But next to the "Back"-Button should be an option to go to the query results again. Helpful to look for articles in native language. Summary and translation is saving time of selecting.

The summarization and the extraction of common elements in the dossier.

General idea of finding, saving content, etc is strong

The summarisation tool is very promising; the translation tool is helpful but quality could be better.

Summarization of lengthy text Translations if they were of better quality

the summarization tool was pretty good. it provided a uniform glimpse on what was to be expected from an article. the related content was shown in German language. I liked it, but I was a bit puzzled about the language.

Relates Articles,

Summaries and summary translations could be hugely useful if they were much better.

Finding relevant asticles in many languages

Perhaps the provision of more in detail insights related to the media monitoring strand (and more precisely regarding influencers' data)

The summarizing of articles.

the sentiment / polarity analysis, while also the transcription service was very interesting

Τα θετικά και αρνητικά κείμενα για ένα θέμα σε σχέση με την κλίμακα.

transcription, sentimentality

better link between the different articles with similar theme

related media, transcriprion and run in depth semantic analysis

In depth semantic analysis

The summarisation tool is very helpful.

The translation in other languages

Portfolio analysis

Text summarization, translation and multimedia analysis

Summarisation

Topic and event detection is very interesting

Potfolio analysis

There are external links to find more about entities and ways to use word cloud and entities to filter articles.

Further remarks?

Entities with full name. Good that you can click on it and it will highlight it in the text. But the functionality is not clear. Not sure what Sentimentality and Polarity is trying to say.
Depth analysis could be embedded in the headline of the article.

The summarization texts need to be shorter. Also, it would help to get the results presented in more readable form - clearer distinction of titles and summaries for example. Or added multimedia elements. The current form is very dense.

Keyword cloud was rather useless

make use of alt-tags. sometimes the system is not self-explanatory. show a loading icon. sometimes loading pages took a while. provide immediate feedback! i was not able to fulfill the advanced search. The system wouldn't let me type anything into the search field. Get rid of the starting screen at the beginning of the evaluation. instead use the general search interface. concerning the questionnaire: nice and short questionnaire. one point: help evaluators in pinpointing them to the portfolio link by saying: press on the portfolio icon in the upper left corner.

Selection of articles appears haphazard, there is no discernable ordering of results by relevance or any other clearly noticeable factor. In fact, a quick search on Google News provides better results more quickly. Entity recognition is wrong most of the time. The same goes for concepts: I selected only articles dedicated to US and UK energy topics, yet Multisensor gives "Germany" as one of the main concepts. Even worse: When I hit "run analysis" on one and the same portfolio several times, Multisensor returns *different* word clouds and concepts every time! Summaries are okay-ish, but translations do more to obfuscate the meaning than help understand it. Multimedia search for "energy policy" was often wildly wrong, returning results referring to asteroids, a Samsung smartphone, or environment without clear relation to energy policy.

Why not to use google search?

none

Extremely helpful, easy to use and very efficient

Database credibility?

Some features such as the entities or the topic cloud I did not understand. They had no additional value. But I do like the overall approach of having one tool to do this kind of research.

Entities with full name. Good that you can click on it and it will highlight it in the text. But the functionality is not clear. Not sure what Sentimentality and Polarity is trying to say.

Depth analysis could be embedded in the headline of the article.

The summarization texts need to be shorter. Also, it would help to get the results presented in more readable form - clearer distinction of titles and summaries for example. Or added multimedia elements. The current form is very dense.

Keyword cloud was rather useless

make use of alt-tags. sometimes the system is not self-explanatory. show a loading icon. sometimes loading pages took a while. provide immediate feedback! i was not able to fulfill the advanced search. The system wouldn't let me type anything into the search field. Get rid of the starting screen at the beginning of the evaluation. instead use the general search interface. concerning the questionnaire: nice and short questionnaire. one point: help evaluators in pinpointing them to the portfolio link by saying: press on the portfolio icon in the upper left corner.

Selection of articles appears haphazard, there is no discernable ordering of results by relevance or any other clearly noticeable factor. In fact, a quick search on Google News provides better results more quickly. Entity recognition is wrong most of the time. The

same goes for concepts: I selected only articles dedicated to US and UK energy topics, yet Multisensor gives "Germany" as one of the main concepts. Even worse: When I hit "run analysis" on one and the same portfolio several times, Multisensor returns *different* word clouds and concepts every time! Summaries are okay-ish, but translations do more to obfuscate the meaning than help understand it. Multimedia search for "energy policy" was often wildly wrong, returning results referring to asteroids, a Samsung smartphone, or environment without clear relation to energy policy.

Why not to use google search?

none

Extremely helpful, easy to use and very efficient

Database credibility?

Some features such as the entities or the topic cloud I did not understand. They had no additional value. But I do like the overall approach of having one tool to do this kind of research.

B. Pilot Use Case 2: Commercial Media Monitoring

B.1. Questionnaire

MULTISENSOR - PUC2: Final Prototype Evaluation Questionnaire (Media Monitoring)

* Obligatory

Please answer all questions in English!

Entry Questions

Name *

Please provide your age: *

- 18-24
- 25-34
- 35-44
- 45-54
- 55+

Please indicate your gender: *

- Female
- Male

Please provide the country you live in: *

Please indicate your native language(s): *

Please provide your current occupation: *

- Editor
- Analyst
- Sales Professional
- Developer
- Other

User Tasks: Assessing MULTISENSOR

Please log into the MULTISENSOR UC2 prototype and select the language for your user profile. Go to the profiles tab. In here, you will find preconfigured searches for continuous analysis projects. Select the search profile "household appliances". In the following we are asking you to answer a few questions on MULTISENSOR and its features. Please use the scale as explained below.

Search Section / Single Results List

Your task is to select several relevant articles for your client, who is a household appliances company. Go to the search section and query for articles from the household appliances domain, e.g. using the term "appliances" or a household appliances manufacturer like Miele or Electrolux. The system will return a result list with single articles from which you can freely select relevant content. Go to the search results ("single results" display). In here, you can identify relevant content for analysis on a basis of individual articles. For this task, use the functionalities provided by the MULTISENSOR system.

Extractive Summarisation: Single Document

Initiate a summary for one or more selected article(s) and compare it to the original text.

The summaries are useful for quickly deciding on the relevance of articles. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

The quality of the summaries is adequate for fulfilling the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Keyword-Based Summarisation

You can tailor your summary to a company or person or other keyword of interest, e.g. your client. Initiate a keyword-based summary for one or more selected article(s) and compare it to the original text.

The summary adequately mirrors the content of the article from the client's point of view. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Translation

In case your search results are not in the same language as your account, you have the option of translating the results into the account language. Please be aware that this service is not aiming for a perfect translation, but is supposed to support you in assessing whether a foreign language article might be relevant.

The translation is useful for quickly selecting relevant content. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Entities Feature

For each article, you can display the entities (i.e. persons, companies and locations) that are contained in the text. Open the entities information for one or more selected article(s).

The list of entities found in the text is useful for quickly selecting relevant content. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Category Feature

All articles are categorised into the following categories: 1) Economy, Business & Finance 2) Science & Technology 3) Lifestyle & Leisure 4) Health 5) Nature & Environment 6) Politics Assess the information for selected articles.

You found the category information to be useful for quickly selecting relevant content. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

6- n/a

Sentiment Feature

Assess the displayed information for sentiment on the work sheet. Sentimentality is rated as the level of arousal or total amount of sentiment (irrespective of whether it is positive or negative) contained in the text, i.e., how emotional is the text? The scale is 1 to 10.

You found the sentiment indicator to be useful for quickly selecting relevant content. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Analysis Section

Go to the analysis section. In here you will find all the relevant content previously prepared and selected. Imagine you are an analyst and want to get a quick overview about the persons, companies and topics mentioned in the articles.

Charts

Assess the displayed information for relevance and correctness on the work sheet.

The charts in the analysis section provide helpful information about the article selection. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Multi-Document Summary

By clicking on the charts you will initiate a multi-document summary. Please select an entity (person or company) with up to 10 articles, click on the bar and compare the resulting summary to the original texts.

You found the provided summary helpful in getting a quick overview of the content of the articles. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Further remarks on the analysis charts or multi-document summarisation?

Influencer Section

Go to the Influencer Section of the prototype. In here, you can find Twitter influencers for the household appliances domain who have been active today. Apart from established meta data, an influencer score is being displayed. The influencer score indicates how influential a user is in a social network, which relates to the number of his/her followers and how often these followers share that user's content. In other words, a user is influential in a social network if his or her activity level has a significant effect on others' activity levels and, consequently, on the site's overall page view volume. The score is not scaled. High scores indicate highly influential users and low scores (close to 0) less influential. Use the visualisations to find today's most influential users and communities and assess the application below.

z

Task Feedback

Please answer a few questions relating to your perceived effectiveness, efficiency and satisfaction when using MULTISENSOR.

Effectiveness

I was able to successfully complete the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

I could complete the task on first attempt. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

I could complete the task without external assistance. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

I did not notice any persistent errors when performing the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Further remarks with regard to effectiveness:

Efficiency

It did not take a lot of time to understand MULTISENSOR and learn about its functionalities.

*

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

The MULTISENSOR system was easy to use and the main functionalities were easy to find. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

It did not take a lot of time to perform the particular task compared to my current method of handling the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

It did not take a lot of time to perform the particular task compared to the use of alternative tools that I have already used. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

Further remarks with regard to efficiency:

Satisfaction

I felt in control when I used MULTISENSOR for performing the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

It made me more productive when I used MULTISENSOR for performing the task. *

1 - Strongly disagree

2 - Disagree

3 - Neutral

4 - Agree

5 - Strongly agree

n/a

The use of MULTISENSOR was overall a satisfying experience. *

- 1 - Strongly disagree
- 2 - Disagree
- 3 - Neutral
- 4 - Agree
- 5 - Strongly agree

n/a

The MULTISENSOR interface is intuitive and easy to use. *

- 1 - Strongly disagree
- 2 - Disagree
- 3 - Neutral
- 4 - Agree
- 5 - Strongly agree

n/a

I would recommend MULTISENSOR to a colleague or friend. *

- 1 - Strongly disagree
- 2 - Disagree
- 3 - Neutral
- 4 - Agree
- 5 - Strongly agree

n/a

Further remarks with regard to satisfaction:

Which functionality/functionalities of the MULTISENSOR system do you consider as most promising and suitable for further development and subsequent exploitation?

B.2. Summary Evaluation Results

Answer	Further remarks on the analysis charts or multi-document summarisation?
1	Highlights are missing. Sources of information in multi-document summaries are not transparent. Charts are not well designed.
2	The summarisation contains irrelevant data like mail addresses or telephone numbers.

3	I experience "Summaries of all selected articles" to be very confusing, where does an article end, where begins the next one? The transition between them is too fluent. Which content originates in which source?
4	The tool summarises articles by extracting sentences and putting together a new text - it ignores languages, so what we get is one sentence in English, one in German, the next one in English that is rather confusing. When reading the summary we do not get introduced to relevant people / entities, they just appear with their last name in the text. Unfortunately, a mixed language summary and losing relevant information is not so helpful.
5	The format of the multi-document summarisation e.g. the lack of paragraphs isn't very user-friendly. It's easier to get a quick overview of the articles by reading the headlines than by reading the summarisation.
6	It is very useful to store the selected documents from other searches to compare them all together.
7	It would be interesting if charts could be used to filter out documents and then new charts are generated.
8	I prefer shorter summaries.

Answer	Further remarks on the influencer features?
1	Charts need a lot more information.
2	Why is the score so tiny? --> 0,082 cent.
3	"Influencers by relevance": Why is the Influencer with the highest score depicted last in the chart? I would expect to be referred to the influencer when I click on the chart, but nothing happens. I don't quite understand the value of this interactive chart. "Influencer meta data" is very clear and easy to understand.
4	Do we really need figures like 0.099 indicating a high relevance? Could that not be 9/10 - relevancy or something comparable?
5	It would have been interesting to have a look at larger networks with more than only one centred user. (Most networks only contained two users.)
6	I would like to see the names in the graph or even better, their profile pictures.
7	The graph does not show why users are related (similar tags, retweet?)
8	It would be great to identify the topic and not only the Twitter.

Answer	Further remarks with regard to effectiveness:
1	A preselection - articles are relevant or irrelevant - would be good.
2	The "lens" next to the search is not clickable, you need to click on the green "GO" to be successful.

3	<p>At first, I searched with too many restrictions on countries and languages and got no results. It took a while to understand that I needed to search broadly. (Sure, the tool will provide more data in the future). It takes time to understand the whole tool, especially the article selection. Analysis & Influencer are easy to comprehend.</p> <p>The search contains many options for the article list. I don't quite understand how the following articles move up when I mark an article. Date and source should receive more emphasis for the reader, I hardly looked at them. It would be good to have a deselection option for "Country" and "Language" in the filters, so I would not have to actively choose. When clicked, I would need to select another country.</p>
4	The keyword-based summary wasn't distinct from the "normal" summary. The keyword was included though, so that is not necessarily wrong.
5	Some functions are not intuitive.

Answer	Further remarks with regard to efficiency:
1	The summaries were too long, reading them was therefore too time-consuming.
2	I cannot really compare the MULTISENSOR to other ways of performing particular tasks. It is too much of a test setup and the tasks I perform are too much of a daily routine and therefore I perform them too fast to compare the time I needed to the time it took me performing the MULTISENSOR test.

Answer	Which functionality/functionalities of the MULTISENSOR system do you consider as most promising and suitable for further development and subsequent exploitation?
1	Search results in different languages; summary.
2	Influencer network. Automatic translation. Entity extraction.
3	Summaries and translations.
4	The influencer tool provides a very good overview (apart from the upmost graphic). The analysis is also comprehensible. I had problems with the search, though.
5	The recognition of entities could be very useful e.g. to capture relevant opinion leaders in a text.
6	Comparison of selected articles, summaries and translation.

Think-Aloud-Notes (Direct Interviews)

User 1:

Entities: These are incomplete and pure first names are not helpful. The functionality itself is a good idea if the quality is good. The entities could be displayed in a more clearly categorised way.

Summary: The beginning is a bit strange, otherwise the summaries are very good. It seems logical to have a summary of a certain percentage of the original article. Lengthening or shortening it would be a nice-to-have.

Keyword-based summary: Text entry did not provide results. Otherwise, having customised summaries comes quite handy.

Metadata: Categories and language detection are helpful for quickly selecting content. Sentimentality would be more interesting when creating analysis reports than during article selection.

Analysis charts: The colours are hard to distinguish. Provided results are helpful if the quality of detected entities is sufficient.

Multi-document summary: This is a nice idea for analysts. However, presently the quality is not good enough. It seems that the system only chooses the first few sentences of every article. I would prefer to have a list display instead of a text block, as the relations between sentences become unclear. This list should contain e.g. the sentences with the highest sentimentality or the most relevant statement otherwise. It should be ranked and offer the possibility to deselect irrelevant information.

Influencers: The influencer score remains unclear to me, I would prefer to rely on the established known metrics to determine relevance.

User 2

Search: Semantic search is impressive, does it take Boolean operators? The results display and selection process is very intuitive. The individual articles could be separated more clearly from each other.

Metadata: In the media monitoring work context, sentiment will be easily confused with tonality. The idea of evaluating sentimentality instead of or in addition to tonality is quite foreign to me. Might however be interesting, even though I don't have a concrete use case for it.

Summary: This feature is nice. For longer texts, I guess I would only want to read the summary.

Keyword-based summary: It is practical to have the option to put focus on a particular keyword.

Multi-document summary: I would not read the multi-doc summary in this format. Instead, I would like it to focus on similarities and differences between the articles and I would prefer a list of bullet points.

Influencers: Having a score for influencers is generally a good idea, but there are many algorithms for this out there. I usually question these scores and would need to critically assess how they are computed. Are the depicted hashtags static? It would be a better product, if hashtags would be extracted out of the results from a generally interesting query. This way the system would not only depict hashtags I already know but be a kind of trend indicator.

Network analysis: This too is nice in principle and can be useful for finding relevant influencers. I would like to see more user metadata in the mouse-over.

User 3:

Search: For the article selection process, it would be better if you could select a default for accept or reject in order to save clicks. The selected articles are instantly stored, but I would need also a method to reverse the process in case I make a mistake.

Summary: This is not bad, but I would like to put more emphasis on the first or first two sentences.

Keyword-based summary: Good idea, but does not seem to work when I enter only a part of a word entered in the text such as "Beratung". (Entering a keyword without clicking on the +button will trigger an error)

Entities: This is useful, but I would need to see the full names of persons. Also, sometimes single characters are shown. Translating the entities into English when having a German profile is a little confusing.

Sentiment: A nice-to have feature. I haven't previously used a score like this.

Multi-document summary: This looks confusing. I guess I would read the original articles.

Influencers: To me, an influence score should not only consider the followers and share, but also how relevant an influencer is with regard to the topic I am interested in, e.g. fashion. Making a qualified statement who is influential for a topic is a very difficult decision. Maybe looking at the profile descriptions could help, or setting a threshold number of relevant posts that needs to be surpassed before being considered an influencer. Also, it would be important to reliably and automatically separate natural persons from companies and bots. Without these two components, I guess I would need to check all of the users that are displayed manually and the score would be of little help.

User 4:

Search: The semantic search is quite cool. In the article display, I miss the highlights for my search terms.

Workflow: Having finished the selection process, I would like to have some kind of success notice or the articles should have disappeared. Summarisation in the article selection process offers no great advantage over having snippets surrounding the search term.

Summary: I see licensing problems for summaries that are completely made up of extracted sentences. In this case, it would be necessary to have a restricted number of characters. Anyway, isn't there a DIN Norm for summaries with regard to their length?

Entities: It would be better if I could kick out incorrect entities. I think it's great they are translated into English; however, you should highlight the entities (and translations of them) in the text.

User 5:

Search: The visualisation of the results is difficult to read as there are neither breaks nor highlights.

Keyword-based summary: The selection process for the keywords and entities is not intuitive. I didn't notice I could click them and needing to click on the add-button instead of pressing enter is annoying.

Multi-document summary: Basically, this is a good idea, but the summary needs more structure. It would be good to focus on the main statements of the articles and always mention the source of the information. I would like to see contrasting and overlapping information visualised. Quotes might also be interesting. Like in the keyword-based summarisation, you should probably have a focus on your client's interests.

Charts: The charts show entities at article level, don't they? This is a good approach, but maybe you could also display the mentions of an entity per article.

Influencers: The network chart is quite useful to find relevant persons, especially compared with the tree-map, which doesn't provide much insight. To see the changes during the day, I would like to have a count of the tweets that are being displayed in the chart. Also, when I click on an influencer, I cannot immediately see why he is considered influential. The relevant post should be directly linked or be displayed otherwise.

User 6:

Overall: I like many of the technologies and find them generally interesting.

User 7:

Search: The search doesn't start if I click on the lens, but I have to press Go! or Enter. I like the status bar for the search, though.

Summary: Some summaries were too long. The sentences did not create too good a reading flow.

Translation: The translation is sometimes funny, but offers a good help if I need to select articles in foreign languages.

Entities: Some entities were not correct, e.g. "Street", "So", "Again". At least once I had a person symbol for "French". In the charts, I would have liked to see locations categorised, e.g. countries and cities.

Categories: The categories are good and helpful, but sometimes too broad, e.g. when relating more to consumers than science.

Influencers: The influence score should use another scale. Presently, the very small numbers create the impression that all values are very close and this is not helpful when I want to find the most important person. Also, the network analysis is nice, but I would like to see who reacted to whom. The present chart only shows a general connection without visualising directions.

User 8:

Search: The semantic search is very good and would be a great help!

Translation: The translation is helpful in the selection process but using this text in the analysis might be fatal.

Keyword-based summary: I like the concept of customisable summaries.

Influencers: The influencer score doesn't use the friendliest range – why don't you go for whole numbers? I would also really like to know how it is calculated. From the followers, I am more impressed by some of the lower ranking accounts. As to the Treemap, I want it to be clickable. Generally, I would want to restrict the displayed followers to topics that interest me, e.g. coffee machines or the like so I can reach out to the most interested influencers.

Multi-document summary: The abstract includes a mix of languages which is impractical, this should be uniform, e.g. in my account language.

C. Pilot Use Case 3: SME Internationalisation

C.1. General Questionnaire

Pilot Use Case 3: SME Internationalisation – Final Evaluation

Part 1: Entry Questions

1. Please provide your age:

- 18-24
- 25-34
- 35-44
- 45-54
- 55+

2. Please indicate your gender:

- Male
- Female

3. Please provide the country you live in:

4. Please provide your native language(s):

5. Please provide your current occupation:

- SME CEO
- SME Export Manager
- Export Freelance
- Other, please specify

6. Please specify the sector you work on or you are specialised on:

- Industrial
- Construction
- Textile
- Agro-alimentary
- Chemical
- Automotive
- Other, please specify

Part 2: MULTISENSOR Professional User Tasks

Scenario

You are an export manager of an SME that wants to start exporting food products to Europe. You need to look into the different countries to identify the market you will target.

Please mark the box that best matches your opinion. Please answer the questions as fully as you feel able to. If not differently specified, the questions below require an answer on a scale from 1 to 5, as specified below.



COUNTRY

- (1) Please select a country.
- (2) Navigate through the Country sections 'Politics', 'Economy', 'Society' and 'Culture'.
- (3) View the information and assess its quality.

7. The country information displayed is easy to visualise and understand.

1 2 3 4 5

Strongly disagree Strongly agree

8. The country indicators displayed are relevant for internationalisation

1 2 3 4 5

Strongly disagree Strongly agree

9. The country information displayed helps to assess the situation of the country

1 2 3 4 5

Strongly disagree Strongly agree

10. Further remarks on country information

SECTOR & PRODUCT – ARTICLES

- (1) At the left column, please click 'Sector Information' and select Food or Beverage. Assess the articles displayed and its relation to your search.
- (2) Then, please click 'Product Information' and select one. Assess the articles displayed and its relation to your search.

11. The product news information is easy to visualise and understand.

1 2 3 4 5

Strongly disagree Strongly agree

12. The articles displayed match adequately your selected product search.

1 2 3 4 5

Strongly disagree Strongly agree

13. The product news stream displayed is relevant in an SME internationalisation context.

1 2 3 4 5

Strongly disagree Strongly agree

PRODUCT – SOCIAL MEDIA

- (1) Please click 'Product Information' and select one in the drop-down menu. Assess the social media analysis that is displayed.

14. The social media analysis is easy to visualise and understand.

1 2 3 4 5

Strongly disagree Strongly agree

15. The social media analysis is relevant for internationalisation

1 2 3 4 5

Strongly disagree Strongly agree

16. Further remarks on the sector and product information available.
-
-

DECISION SUPPORT – TABLE OF INDICATORS

(1) Please click "Assessment" under "Internationalisation support" and follow the steps indicated in the website.

15. The table of indicators displayed is easy to visualise and understand.

1 2 3 4 5

Strongly disagree Strongly agree

16. The table of indicators displayed is relevant in an SME internationalisation context.

1 2 3 4 5

Strongly disagree Strongly agree

17. The table of indicators helps you to compare and understand differences between countries.

1 2 3 4 5

Strongly disagree Strongly agree

DECISION SUPPORT – RESULTS

18. The Decision Support results help you to compare and identify possible markets to export to.

1 2 3 4 5

Strongly disagree Strongly agree

19. The Decision Support tool is relevant in an SME internationalisation context.

1 2 3 4 5

Strongly disagree Strongly agree**20. Further remarks on the Decision Support tool**

Part 3: General questions about MULTISENSOR and its Tasks**Effectiveness****19.** I was able to successfully complete the tasks.

1 2 3 4 5

Strongly disagree Strongly agree**20.** I could complete the tasks on first attempt.

1 2 3 4 5

Strongly disagree Strongly agree**21.** I could complete the tasks without external assistance.

1 2 3 4 5

Strongly disagree Strongly agree**22.** I did not notice any persistent errors when performing the tasks.

1 2 3 4 5

Strongly disagree Strongly agree**Efficiency****23.** It did not take a lot of time to understand MULTISENSOR and learn about its functionalities.

1 2 3 4 5

Strongly disagree Strongly agree

24. The MULTISENSOR system was easy to use and the main functionalities were easy to find.

1 2 3 4 5

Strongly disagree Strongly agree

25. Using MULTISENSOR I would save time when doing the initial steps to identify countries for internationalisation.

1 2 3 4 5

Strongly disagree Strongly agree

Satisfaction

26. I felt in control when I used MULTISENSOR for performing the task.

1 2 3 4 5

Strongly disagree Strongly agree

27. It made me more productive when I used MULTISENSOR for performing the task.

1 2 3 4 5

Strongly disagree Strongly agree

28. The use of MULTISENSOR was overall a satisfying experience.

1 2 3 4 5

Strongly disagree Strongly agree

29. The MULTISENSOR interface is intuitive and easy to use.

1 2 3 4 5

Strongly disagree Strongly agree

30. I would recommend MULTISENSOR to a colleague or friend.

1 2 3 4 5

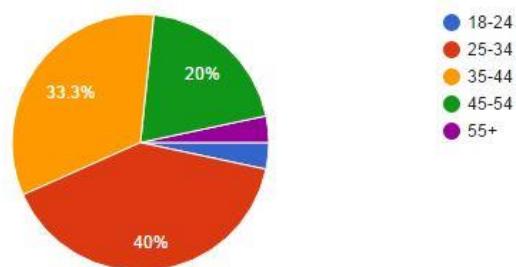
Strongly disagree Strongly agree

31. Please fill in any other remarks or suggestions you may have

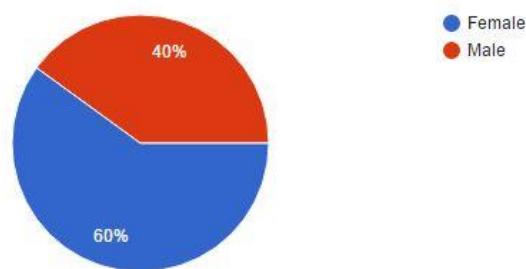
Using MULTISENSOR I would save time when doing the initial steps to identify countries for internationalisation.

C.2. Evaluation Answers Summary

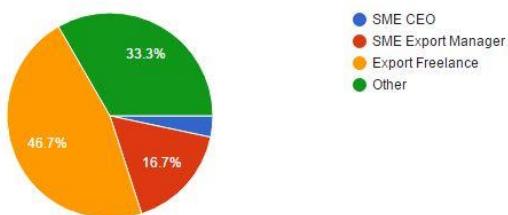
Please provide your age (30 responses)



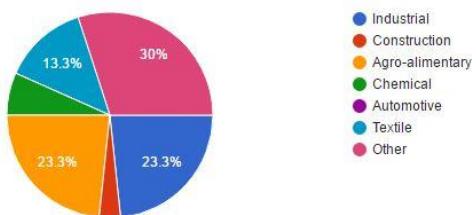
Please indicate your gender (30 responses)



Please provide your current occupation (30 responses)

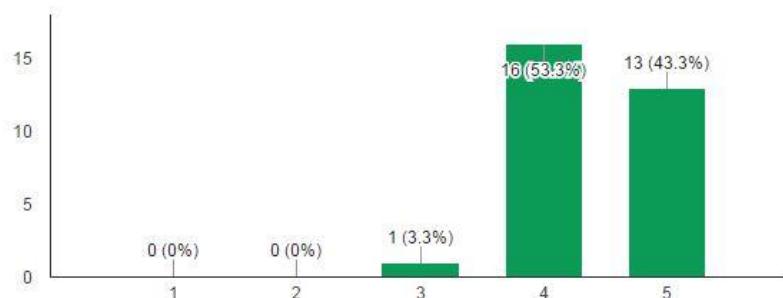


Please specify the sector you work on or you are specialised on (30 responses)

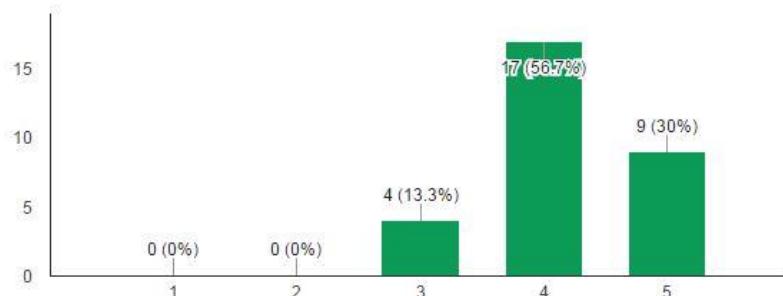


COUNTRY

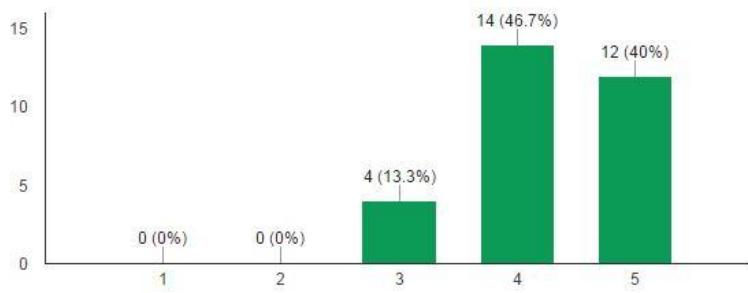
8. The country information displayed is easy to visualise and understand.
(30 responses)



9. The country indicators displayed are relevant for internationalisation
(30 responses)



10. The country information displayed helps to assess the situation of the country
(30 responses)



Country information comments

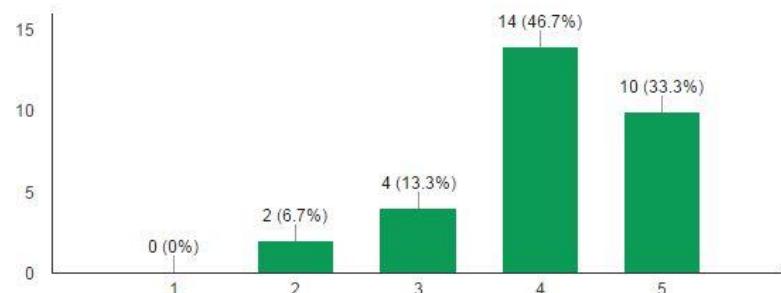
Would be useful to have some charts with the main export sector/goods, and to which countries is exporting most and same with the imports.

There are quite macro indicators which are good for a first evaluation although it is difficult to extrapolate it for all products.

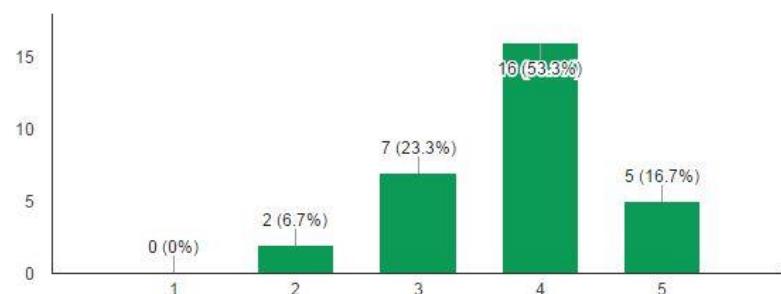
Where the indicators chart, I would suggest to indicate the average number of each topic to get to compare and know where the country is positioned.

NEWS & ARTICLES

12. The product news information is easy to visualise and understand.
(30 responses)

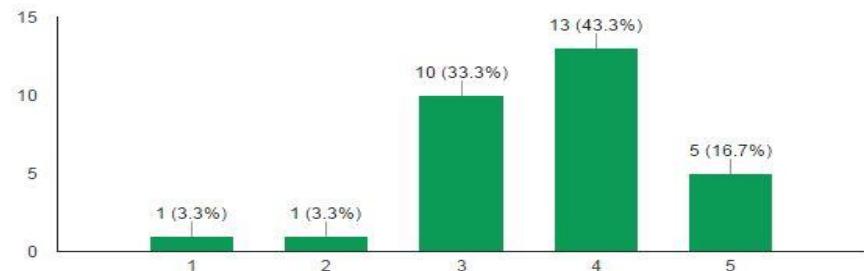


13. The articles displayed match adequately your selected product search.
(30 responses)



14. The product news stream displayed is relevant in an SME internationalisation context.

(30 responses)



News & Articles comments

In the articles displayed I miss information about the market trends, more commercial aspects and for the social media is not really clear what you are looking for.

The articles do not adjust much at market information, they are more general. The social media part interested me as it some companies can need it and it is more difficult to find information on it.

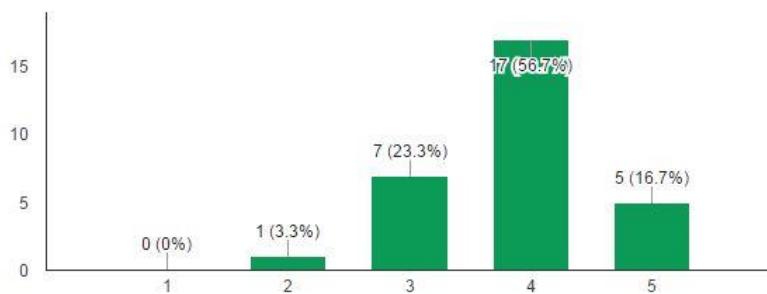
I found the news very fresh and interesting although some articles were not entirely related to markets. I believe the social media analysis is a very good point.

Information could be filtered by country.

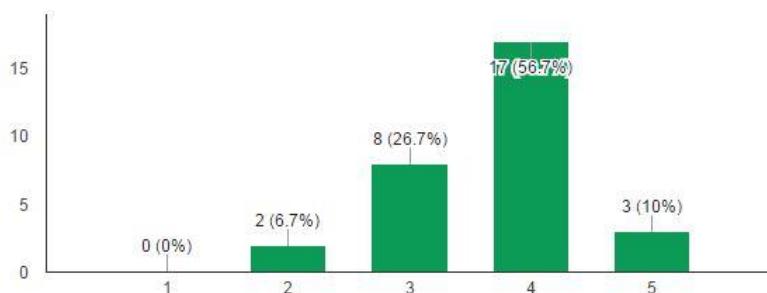
Very useful and interesting.

PRODUCT - SOCIAL MEDIA

15. The social media analysis is easy to visualise and understand. (30 responses)



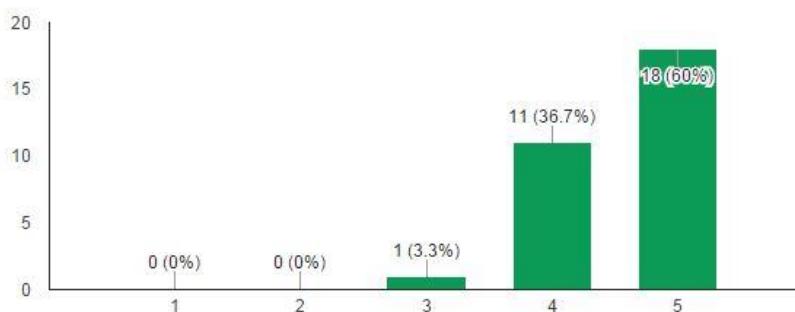
16. The social media analysis is relevant for internationalisation (30 responses)



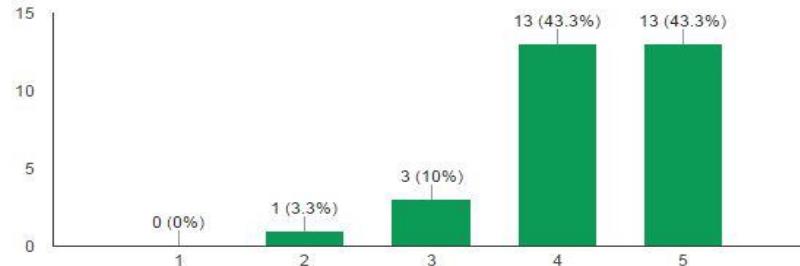
DECISION SUPPORT

TABLE OF INDICATORS

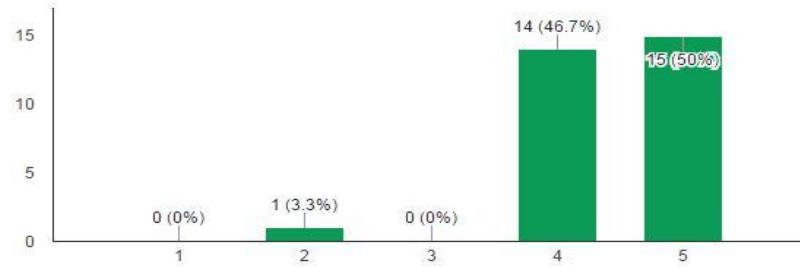
18. The table of indicators displayed is easy to visualise and understand.
(30 responses)



19. The table of indicators displayed is relevant in an SME internationalisation context
(30 responses)



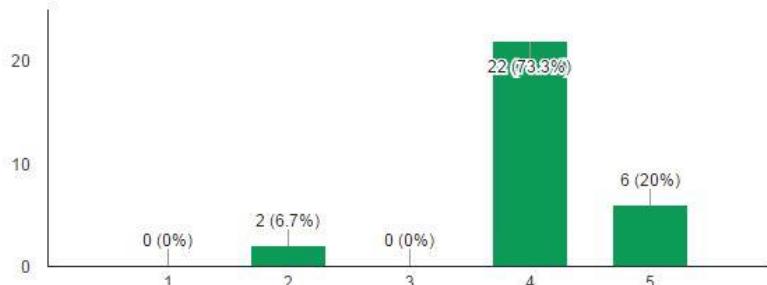
20. The table of indicators helps you to compare and understand differences between countries.
(30 responses)



RESULTS

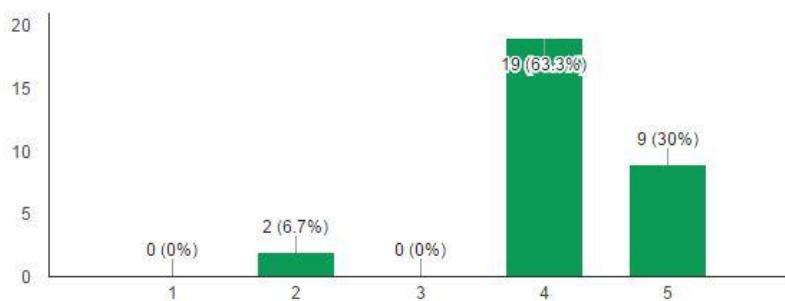
21. The Decision Support results help you to compare and identify possible markets to export to.

(30 responses)



22. Decision Support tool is relevant in an SME internationalisation context.

(30 responses)

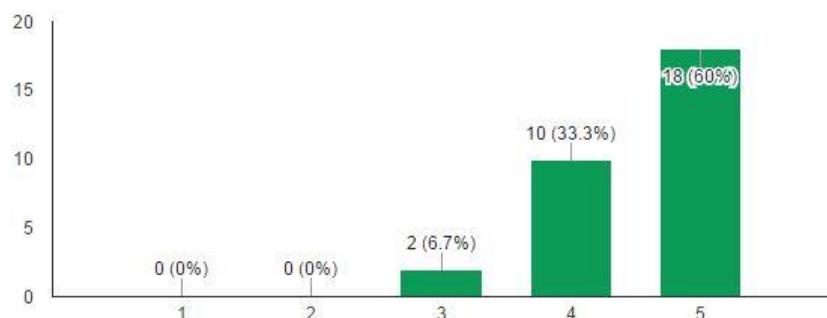
**Decision Support comments**

I like the fact that you can compare countries.

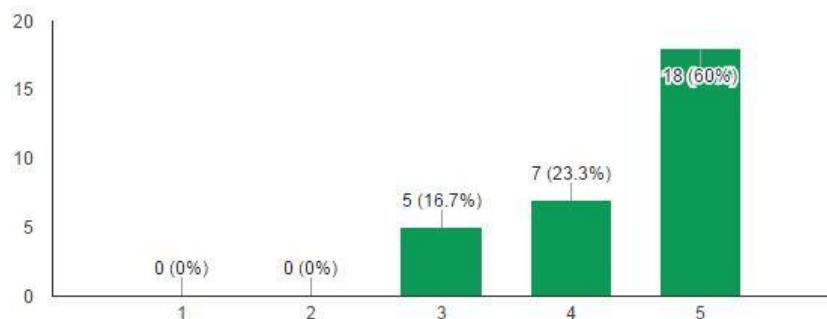
Explanation for the suggested countries would be good.

GENERAL QUESTIONS

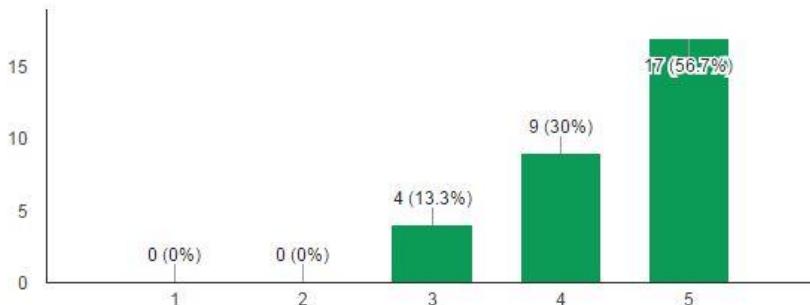
24. I was able to successfully complete the tasks. (30 responses)



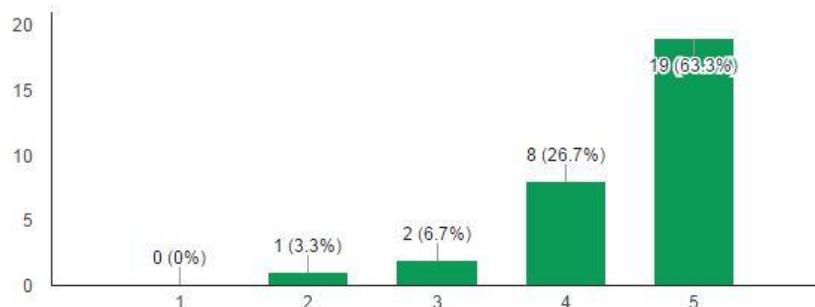
25. I could complete the tasks on first attempt. (30 responses)



26. I could complete the tasks without external assistance. (30 responses)

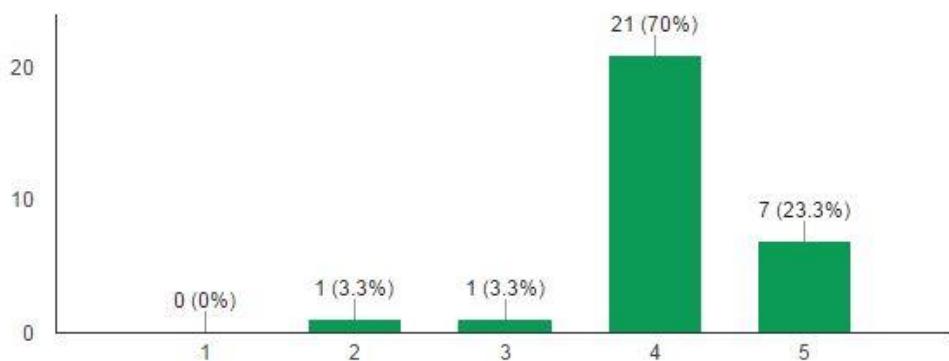


27. I did not notice any persistent errors when performing the tasks.
(30 responses)



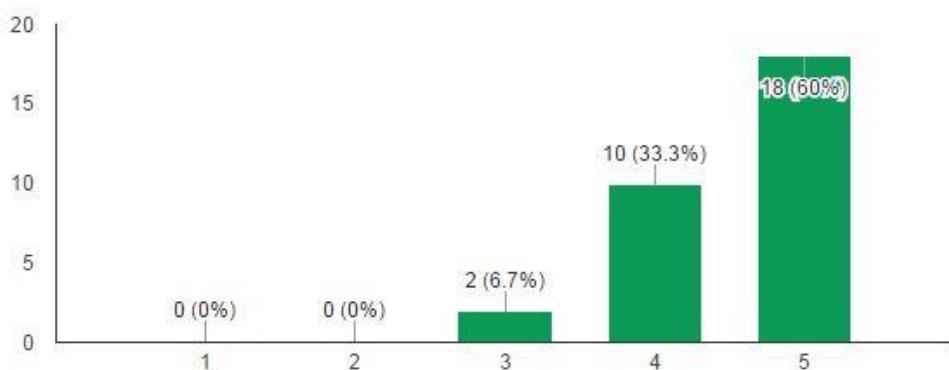
29. It did not take a lot of time to understand MULTISENSOR and learn about its functionalities.

(30 responses)



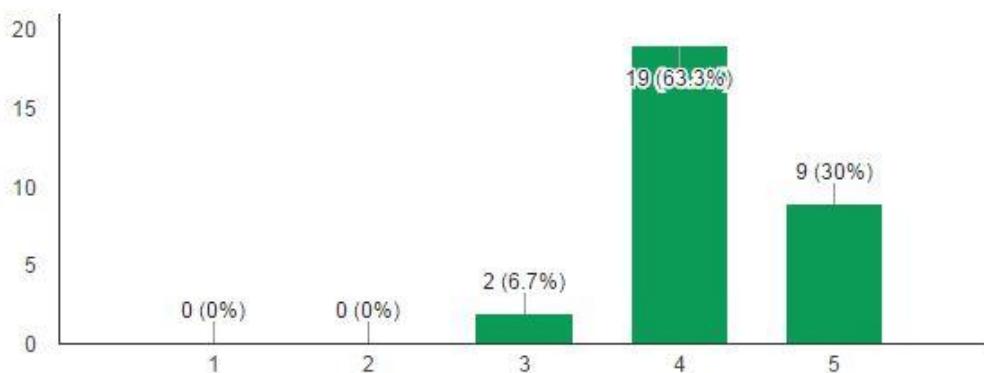
30. The MULTISENSOR system was easy to use and the main functionalities were easy to find.

(30 responses)



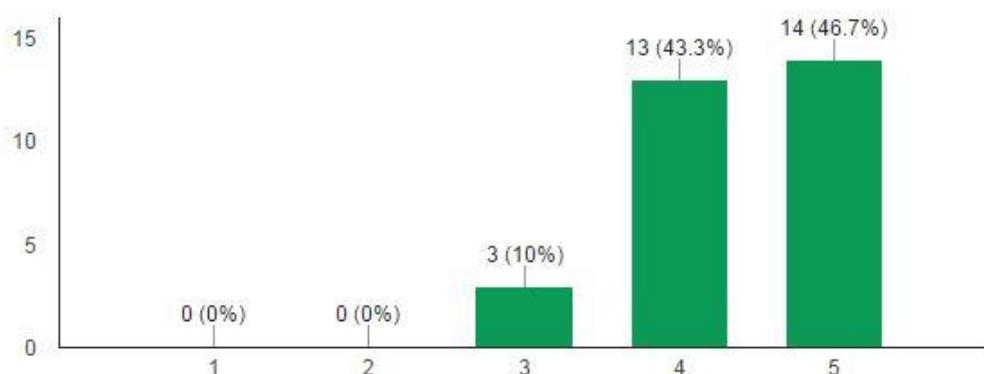
31. Using MULTISENSOR I would save time when doing the initial steps to identify countries for internationalisation.

(30 responses)



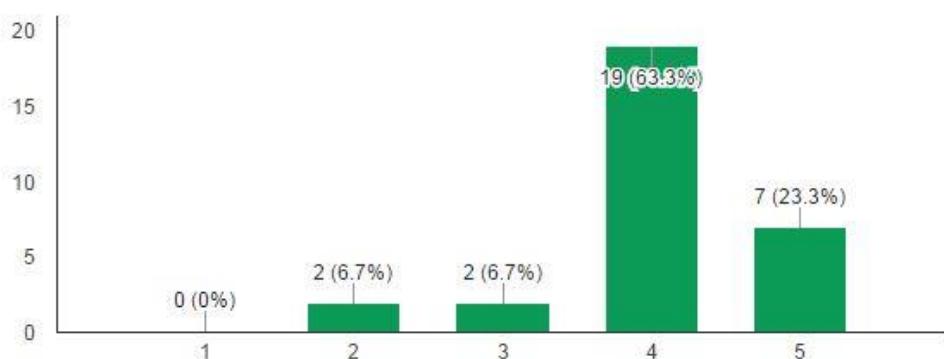
32. I felt in control when I used MULTISENSOR for performing the task.

(30 responses)

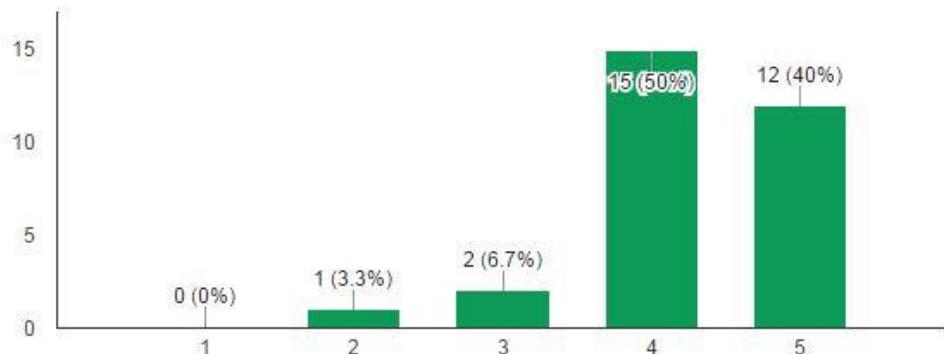


33. It made me more productive when I used MULTISENSOR for performing the task.

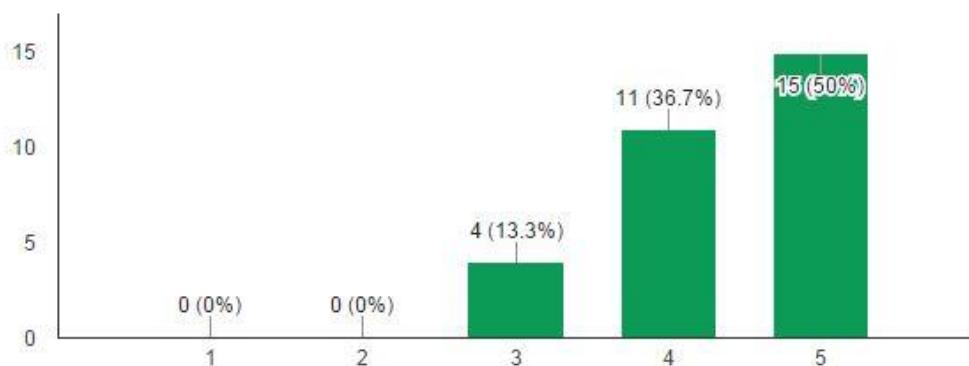
(30 responses)



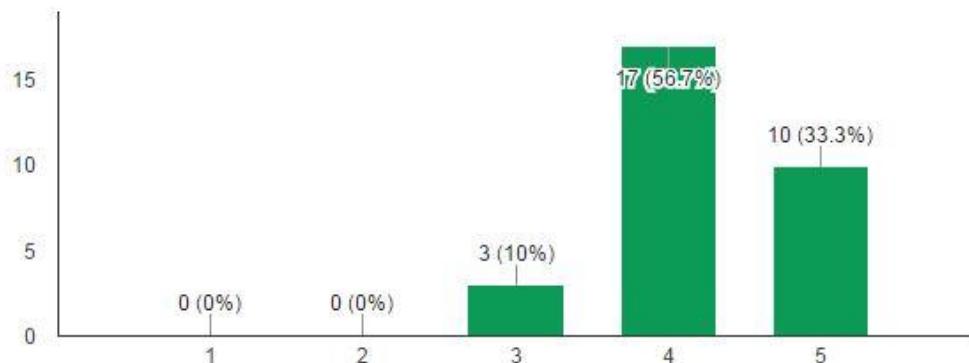
34. The use of MULTISENSOR was overall a satisfying experience (30 responses)



35. The MULTISENSOR interface is intuitive and easy to use. (30 responses)



36. I would recommend MULTISENSOR to a colleague or friend. (30 responses)



Exploitability question answers

It's easy to find information.

Above all, I consider that the section of "Assessment" gives a good view in order to start export processes. A part from that, the other sections help to develop an initial process.

Social Media

The assessment part in the internationalization support.

Decision support can have some options. But information on legal, tax issues etc are also to be considered...

News could be one if very related to market, economy, competitors etc. Decision support is helpful too

Social media analysis.

Different information in one place

This is a good initial assessment to enter a country. Although I believe having more information on the following subjects could be relevant: - distribution channels of the product -consumer preferences / use - Basic legal advice

I like the social media part, especially the fact that you can find influencers.

The system to compare the markets.

Internationalisation support.

Database information per country.

Graphics and decision support.

Comparing countries.

Social media.

The social media and the assessment in order to compare two countries.

Information about countries, if it is updated.

Country comparison is a good tool to rapidly assess markets and see different combinations and alternatives.

I liked the decision support but having more products would be better. also explaining a bit the decision system that it is used.

overall country information is useful to have it in one website and possibility of comparing.

the assessment table is useful and the suggestion of the countries. maybe putting also the averages would help to see the situation of the countries.

decision support can be good for companies that assess others

D. Hardware Infrastructure

EVERIS Server

The server runs Ubuntu Linux 14.04.1 LTS (“Trusty”) on x64 architecture. Ubuntu is hugely popular and as such, Personal Package Archives (PPAs) and vendor repositories are readily available providing very recent versions of core packages of MULTISENSOR (mongodb, elasticsearch, nodejs, maven, nginx).

The main server, called **msgrender1**, is hosting the Content Extraction Pipeline Services, the repositories and the three Use Case applications.

The server has the following specifications:

- 16x x64 core (52 ECUs)
- 122 GB RAM
- 300 GB local SSD storage (xfs)
- 100 GB EBS SSD storage (ext4)

ONTOTEXT Server

All of Ontotext services are deployed on two virtual hosts, which are part of a bigger physical machine. Below is the list of deployed services and a detailed description of the hardware infrastructure:

- GraphDB
- RDF Storing Service
- RDF Validation Service
- ElasticSearch engine
- Bulgarian Dependency Extractor
- News On the Web (NOW) platform
- NOW pipelines

LINGUATEC Server

The Linguatec Services are hosted on their own hardware by the professional housing provider “QSC” in Munich. Corresponding to the requirements of each service, we have used the following hardware infrastructure based on one HP Blade Centre C7000 and 6 IBM3550 Servers:

Named Entities Recognition

The Language Identification Module is not a resource intensive component, but it depends on the number of simultaneously requests in runtime.

Web Server.

An apache web server on Ubuntu Linux OS on a virtual server with reserved resources:

- CPU: 2 x Intel Core 2 (2.66 Ghz, 128K cache)
- RAM: 4GB
- HDD space: 30GB

Java Applications.

An Ubuntu OS on a dedicated server with following hardware characteristics:

- CPU: Intel Dual Core 4 (2.66 Ghz, 128K cache)
- RAM: 16GB
- HDD space: 100GB

Language Identification

The Language Identification Module is not a resource intensive component, but it depends on the number of simultaneously requests in runtime.

Web Server.

An apache web server on Ubuntu OS on a virtual server with reserved resources:

- CPU: 2 x Intel Core 2 (2.66 Ghz, 128K cache)
- RAM: 4GB
- HDD space: 30GB

Java Applications.

An Ubuntu OS on a virtual server with reserved resources:

- CPU: Intel Dual Core 4 (2.66 Ghz, 128K cache)
- RAM: 16GB
- HDD space: 1000GB

Machine Translation Component

The Machine Translation is a resource intensive component depending on number of simultaneously requests in runtime and amount of monolingual and bilingual corpora in preparation time.

Web Server.

An apache web server on Ubuntu OS on a virtual server with reserved resources:

- CPU: 2 x Intel Core 2 (2.66 Ghz, 128K cache)
- RAM: 4GB
- HDD space: 30GB

Database Instance.

A MySQL DB on Ubuntu OS on a virtual server with reserved resources:

- CPU: 2 x Intel Core 2 (2.66 Ghz, 128K cache)
- RAM: 8GB
- HDD space: 2000GB

Java Applications.

An Ubuntu OS on a three virtual servers with reserved resources:

- CPU: Intel Dual Core 8 (2.66 Ghz, 128K cache)
- RAM: 64GB
- HDD space: 200GB

Translation Engine

An Ubuntu OS on a two virtual server with reserved resources:

- CPU: Intel Dual Core 8 (2.66 Ghz, 128K cache)
- RAM: 64GB
- HDD space: 2000GB

For the developer workplace, especially for calculation of language model for the translation engine we use 2 dedicated strong machines with following hardware characteristics:

- CPU: 2 x Intel Core 12 (3.0 Ghz, 128K cache)
- RAM: 2048 GB
- HDD space: 20000GB

For the staging platform we us all components on an compact architecture with following characteristics:

- CPU: 2 x Intel Core 4 (2.66 Ghz, 128K cache)
- RAM: 256GB
- HDD space: 10000GB

Audio Transcription

The Audio Transcription is a resource intensive component. The final system runs on the following hardware:

Web Server.

An apache web server on Ubuntu OS on a dedicated server with following characteristics:

- CPU: 2 x Intel Core 4 (2.66 Ghz, 128K cache)
- RAM: 4GB
- HDD space: 30GB

Database Instance.

A MySQL DB on Ubuntu OS on a virtual server with reserved resources:

- CPU: 2 x Intel Core 2 (2.66 Ghz, 128K cache)

-
- RAM: 8GB
 - HDD space: 2000GB

Recognition Engines:

An Ubuntu OS on four virtual servers with reserved resources:

- CPU: 2 x Intel Core 4 (2.66 Ghz, 128K cache)
- RAM: 128GB
- HDD space: 300GB

For the developer workplace, especially for the calculation of language model and acoustic models we use a dedicated machine with the following hardware characteristics:

- CPU: 2 x Intel Core 12 (3.0 Ghz, as much as possible cache)
- RAM: 2048 GB
- HDD space: 20000GB

PRESSRELATIONS Crawler

Global architecture

pressrelations GmbH hosts its hardware at the Interxion data center in Düsseldorf.

Hardware from Watchguard (security), Kemp (loadmaster) and HP/HPE (server/storages/network) is used. Windows, Linux and VMWare servers run on HP/HPE ProLiant servers, data storages are hyperredundant SAN servers by LeftHand/StoreVirtual.

1st generation crawlers

The 1st generation crawlers at pressrelations is used to find new articles, download and extract textual content and assigning articles to clients.

This process is executed on 20 Windows virtual machines in a redundant system architecture on two HP DL360 G7 and two DL385 G7 physical servers.

The database cluster is located on two DL360p G8 servers.

Data volumes are on:

- 6x HP P4500G2 à 6,5 TB (tier 2)
- 3x HP P4330 à 3,3 TB (tier 1)
- 3x VSA à 10 TB (tier 3)

2nd generation crawlers

The 2nd generation crawlers are linux based using Elasticsearch and Docker technologies.

Storage is located on SAN servers (as described in 1.4.1).

Data indexing is done on four virtual machines using Elasticsearch on two HP ProLiant DL380 G9 servers:

- 2 Intel E5-2620 CPU
- 256 GB RAM

Scheduling, download of textual content and text extraction is done on 24 Docker Containers on 6 physical servers:

- HP DL360 G9 E5-2630
- 256 GB RAM

CERTH Twitter Crawler

The Twitter crawler comes from the previous SocialSensor (FP7-287975) project, in which Stream Manager was developed. Stream Manager contains a number of APIs that collect incoming content relevant to a keyword, a user or a location from a set of social streams (Twitter, Facebook, Instagram, etc.). The Twitter crawler specifically gathers Twitter posts for a set of hashtags, which are pre-specified for each Use Case separately. These posts, as well as information regarding the author and the associations found within the posts are then stored into a MongoDB database. Finally ,the Twitter posts gathered by the Twitter crawler and stored into MongoDB are fed as input to the services (Influential User Detection and Community Detection) of the Social Media Analysis Pipeline (SMAP).

The server, on which the Twitter crawler is located and the Twitter posts are being stored, has the following specifications:

- CPU: 2 x Intel Xeon E5-2620v3 6-Core (2.40GHz 15MB).
- Memory: 128GB (8 x 16GB) PC4-17000P-R 2133MHz RDIMM.
- Storage: 2 x Samsung Pro 1TB.
- Storage controller: Smart Array P440ar/2GB.
- Power supply:2 x 500Wredundant power supply units.
- Server management: HP Integrated Lights-Out 4 (iLO 4).
- Network: 4 x Gigabit Ethernet.

CERTH Server

The services that have been developed by CERTH (Similarity search, Category classification, etc.) are stored and deployed in the CERTH server. The server has the following specifications:

- CPU: 4 cores, Intel i7-4790K CPU @ 4.00GHz
- RAM: 32GB RAM
- 500GB SSD
- 2TB HDD