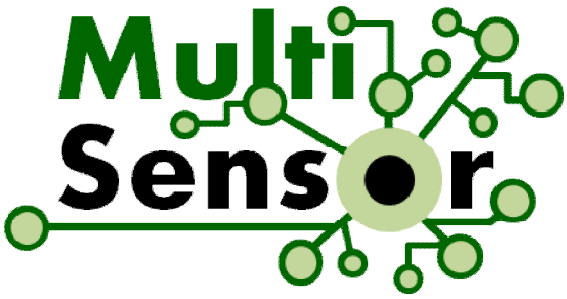


MULTISENSOR

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

FP7-610411

D8.2

User requirements, specification of pilot use cases and validation plan

Dissemination level:	Public
Contractual date of delivery:	Month 9, 31.07.2014
Actual date of delivery:	Month 9, 29.07.2014
Workpackage:	WP8 Use Cases and evaluation
Task:	T8.1 User requirements and content provision T8.2 Specification of the two pilot use cases T8.5 User validation and prototype evaluation
Type:	Report
Approval Status:	Final Draft
Version:	1.1
Number of pages:	101
Filename:	D8.2_RequirementsUseCasesValidationplan_2014-07-29_v1.1.pdf

Abstract

This document details the user requirements for the end-user groups targeted by the project: Journalists and commercial media monitors as well as SMEs (and their managers) that intend to expand into foreign markets. It shows how these target groups depend on information from very heterogeneous and multilingual sources. And it also shows how MULTISENSOR could support the decision making process by identifying, analysing, summarising and contextualising content that is relevant for the respective use case scenario.

co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	06/03/2014	Document initiation	N. Heise (DW)
0.2	24/04/2014	Adding chapters	N. Heise (DW)
0.3	13/05/2014	Review & formatting	T. Wagner (DW)
0.4	03/06/2014	New version after consortium feedback	N. Heise (DW)
0.5	26/06/2014	First consolidated version	N. Heise (DW)
0.6	4/07/2014	Version ready for external review	N. Heise, T. Wagner (DW)
0.7	10/07/2014	External review	Maite Melero (UPF-EUMSSI)
0.8	14/07/2014	Updated version after external review	N. Heise, T Wagner (DW)
0.9	20/07/2014	Internal review	J. de Bruin (PR)
1.0	27/07/2014	Final version	T. Wagner, N. Heise (DW)
1.1	27/07/2014	Final updated version correcting minor errors	T. Wagner, N. Heise (DW)

Author list

Organisation	Name	Contact Information
DW	Nico Heise	nicolaus.heise@dw.de
DW	Tilman Wagner	tilman.wagner@dw.de
PR	Mirja Eckhoff	mirja.eckhoff@pressrelations.de
PR	Michaela Menken	Michaela.menken@pressrelations.de
PR	Romina Gersuni	romina.gersuni@pressrelations.de
PIMEC	Sara Pedraz	export@pimec.org
PIMEC	Teresa Forrelat	Tforrellat@pimec.org
CERTH	Stefanos Vrochidis	stefanos@iti.gr

Executive Summary

The combination of digital technologies, the Internet, social networks and traditional sources have increased and accelerated the flow of information in an unprecedented way. Knowledge workers from every domain are challenged by more and faster communication than ever before. For politics, news and business, these new developments are shaping our world. But how can professionals find and process information that comes from heterogeneous multilingual and multicultural sources? How can they understand information and data that are often ambiguous and noisy in a timely manner and utilize them as basis for their decisions? This is the key challenge for the MULTISENSOR project.

This deliverable explores the requirements and potential solutions through two separate pilot use cases: media monitoring (by journalists and commercial media monitors) and SME internationalisation. It describes the needs of each use case through specific scenarios along with the research methodologies that have determined our approach. The document includes a competitive analysis of the tools currently being used and then sets out the specific requirements for a MULTISENSOR prototype.

The research points towards some clear professional needs around media monitoring and information gathering, which are not met by existing tools. It identifies (amongst others) four main areas where value could be created by the MULTISENSOR project:

- Finding, filtering and pre-selecting information according to relevancy.
- Analysing textual online data with regard to its specific meaning independent of the language in use.
- Automatically summarising the content of complex and heterogeneous data.
- Contextualising data and automatically enriching it with other related content.

The ultimate goal is to create a valid and understandable set of information that can serve as a basis for decision making by professionals such as journalists, commercial media monitors and business managers.

This deliverable focuses on the user perspective and is supposed to ensure the overall user-centric approach of all R&D activities within the MULTISENSOR project. Although some of the expected functionalities are described in great detail, D8.2 does not aim for demonstrating system or technical requirements. This will happen within D7.2 (Technical requirements and architectural design) that will be written based on the user requirements in D8.2.

Abbreviations and Acronyms

ARD	Association of public service broadcasters in Germany
BMCO	Broadcast Mobile Convergence
DAML	DARPA Agent Mark-up Language
DID	Digital Item Definition
DII	Digital Item Identification
DRM	Digital Rights Management
EBU	European Broadcast Union
ETSI	European Telecommunications Standards Institute
IEEE	Institute of Electrical and Electronics Engineers
IP	Integrated Project
IPTC	International Press Telecommunications Council
IST	Information Society Technologies
JPEG	Joint Photographic Experts Group
MAF	Multimedia Application Format
MPEG	Moving Picture Experts Group
NITF	News Industry Text Format
NoE	Network of Excellence
OWL	Ontology Web Language
OWL-QL	Ontology Web Language Query Language
OWL-DL	Ontology Web Language Description Language
RDF	Resource Definition Framework
RSS	Really Simple Syndication
STREP	Specific Targeted Research Projects
SaaS	Software As A Service
W3C	World Wide Web Consortium
XML	eXtensible Mark-up Language
SWRL	Semantic Web Rule Language

Table of Contents

1 INTRODUCTION

The increasing quantity and accelerating flow of information from online and traditional sources combined with the availability of large data sets make it practically impossible to process, analyse and cluster information and data in an adequate time frame without the support of machines and software. Organisations and people that depend on accurate and up-to-date information (e.g. journalists, media monitors or business executives) are looking for systems that streamline information from different sources, cluster what belongs together and discard what is irrelevant in a specific context. Ideally, these systems should automatically understand and summarise the very specific meaning of a specific content item in its specific context regardless of the language in use and the different cultural backgrounds. Research and development within the MULTISENSOR project intends to offer solutions to these needs. In order to ensure that work within the project is guided by a user-centric view and in order to grant general usability and exploitability of the project results, MULTISENSOR has defined two pilot use cases with several scenarios. The three user partners Deutsche Welle, pressrelations and PIMEC have used their specific experience and expertise to identify and define user requirements. The process of requirements gathering, the respective analysis and the requirements themselves will be presented in this deliverable.

1.1 Pilot use cases and scenarios

MULTISENSOR has established two pilot use cases:

(1) **International media monitoring (PUC1)** and
(2) **SME internationalisation (PUC2)**

PUC 1 covers the fact that journalists (scenario 1) as well as commercial media monitoring portals and their clients (scenario 2) are interested in and at the same time challenged by the sheer amount of heterogeneous data streams coming from known or unknown sources. PUC 2 on the other hand reflects a situation where business decisions need to consider a variety of market information from multiple angles. Small or medium-sized companies (SMEs) that intend to expand to foreign markets are particularly challenged by not knowing the relevant sources of information or the local language. This refers to a general market analysis (scenario 1) as well as to the very specific assessment of statutory issues and commercial risks (scenario 2).

1.2 Commonalities and differences

Both use cases and all four scenarios reflect the challenge of having to deal with a large amount of heterogeneous data and information from many different sources in many different languages. But there are also very significant differences: Journalists and commercial media monitors are interested in continuously monitoring a specific topic, brand or campaign. The ultimate goal is to update their audience or clients on a regular basis. Although this regular information will ultimately lead to specific (individual or business) decisions, information gathering and analysis is the priority.

In contrast, the SME internationalisation use case aims for decision support mainly. In order to make a well-founded decision about whether a company should move into a new market, owners and managers need to know about all relevant market indicators, opportunities and barriers. This includes information about consumption, competition, socio-economic indicators as well as information about legal restrictions or other statutory requirements. In fact, the two scenarios of PUC 2 describe two facets of one research task: To identify every piece of information about a potential new market that is relevant for the decision whether a company should enter into that market or not.

1.3 Document structure

These commonalities and differences are reflected by the document's structure. Although both scenarios of PUC 1 deal with monitoring, their underlying concepts and goals are very different. Consequently, the journalistic scenario and the commercial media monitoring scenario as well as their respective requirements will be described and analysed in two different sections (section 2 and section 3). The two scenarios of PUC 2 on the other hand are just two sides of the same coin: Market analysis and statutory analysis both contribute to the one decision whether a company should expand to a new market. Consequently, both scenarios will be discussed within one section (section 4).

According to the description of work, the deliverable is supposed to specify the scenarios within each pilot use case. For each of them, the deliverable should provide:

- the list of target users;
- the context(s), in which the target users solicit decision support and/or summarised information;
- the types of decision support and types of information that must be provided by the system;
- an outline (e.g., in terms of flow charts) how the decision support/information will be ensured;
- the user interaction that will take place at each use case;
- an assessment of the (short, mid and long term) effects of the activities.

As D8.2 is a core deliverable that will ensure the overall user-centric approach of the project, we have chosen - in cooperation with technical partners - a structure that is easy to read and understand. It was considered the best way to bring across what users expect from the MULTISENSOR system by creating a table that names the main requirement by a generic label called "First level requirement" (e.g. 'automatic summarisation' or 'named entities extraction'), briefly describes this first level requirement and then further specifies it by enlisting sub-requirements that we call "Second level requirements". We have also prioritised the first level requirements according to the findings from the requirements elicitation process. Although some of these first and second level requirements are described in great detail, D8.2 does not aim for demonstrating technical or system requirements. This will be done within D7.2 (Technical requirements and architectural design) that will be written following D8.2.

The individual use case sections follow a common structure:

1. **Targeted user groups and usage context:** describing the main user groups and the specific work context
2. **Specific topic:** describing a specific and very practical topic that is suitable for illustrating how MULTISENSOR could contribute in a day-to-day working environment.
3. **Elicitation of user requirements:** describing the methodology and the process of identifying user requirements (including some preliminary findings).
4. **Core user requirements:** the already mentioned table naming and describing first and second level requirements.
5. **Summary and priorities:** summarising the results and prioritising them.

With regard to the description of *user interaction* we have decided to provide preliminary mock-ups of a possible user interface. These mock-ups vary depending on the different use case scenarios and are presented in section 5.

1.4 Common methodology

Task 8.1 is to analyse user requirements with regard to the specific use case scenarios. Although the scenarios are different, we have decided to follow a more or less similar approach. First of all we have analysed the current market situation and what kind of solution existing tools are offering for the targeted user groups. Based on the analysis of this market survey, we have carried out interviews and discussions with relevant stakeholders that have been chosen individually by each user partner with regard to the specifics of their individual use case scenarios. The format of these interviews and discussions (one-to-one interviews, focus groups and committees) did vary depending on the provisions of the DoW and depending on what was considered the most suitable approach. Details of the respective methodology will be presented under the specific use case sections.

2 PILOT USE CASE 1 (SCENARIO 1): JOURNALISM

2.1 Targeted user groups and usage context

Over the last couple of years media companies have embraced and integrated social media into their workflows and their output. Most journalists have social media accounts where they engage directly with audiences and distribute their content, but news organisations are also interested in the strategic use of social media as part of their wider newsgathering and audience engagement plans. Furthermore, the increasing availability of large data sets has led to new phenomena such as data journalism where storytelling is based on a thorough analysis of "big data". Our research in newsrooms reveals three distinct roles, which have guided our requirements.[1]

- **Editors and managers:**
  Supervising a wide range of the production processes within news outlets (sections of newspapers, a television or radio programme), these senior figures are interested in spotting stories or angles that rivals may have missed. They also need to be able to find and understand the level of audience interest in a particular story at any point in time. Senior managers, brand managers and audience research teams will also be interested in the strategic use of data journalism, collaborative journalism and crowd sourcing in meeting wider business objectives.
- **Reporters and producers:**
  Being at the heart of the process of gathering and communicating the news in text, video or audio (via print, radio, TV or online), this group is very interested in spotting new stories. This may include whole new topics or just angels that rivals may have missed, normally within a specific area of interest, sometimes called a journalistic beat. Nowadays a lot of this work is done via social media, engaging/including the crowd (especially influencers and contacts) in finding quotes for the stories. Others have - as "data journalists" - specialised in the analysis of large data sets. They follow a process of *reverse research* by processing these data sets in multiple ways in order to detect possible patterns that are worth telling to a broader audience.
- **Social media specialists:**
  Since the role of social media has risen immensely within the news gathering process new roles and procedures have arisen in news organisations. There are now departments in charge of combing through social media for relevant information and stories. Identifying trends on a broader organisational level, these departments also inject ideas into the wider editorial process. It is their responsibility to proactively seek and verify eyewitness material for publication when news breaks. Working in teams with reporters or correspondents they support the illustration of wider

[1] The descriptions are based on existing role definitions within Deutsche Welle. They were first introduced this way in Deliverable 7.1 – Use Case & Requirements (internal; authored among others by DW staff) and in Deliverable 9.3 – Market Analysis (public) of the Social Sensor Project (www.socialsensor.eu). The roles have proven to be valid and fitting for a news use case scenario. With a very similar setting in both MULTISENSOR and Social Sensor the roles where transferred and slightly adjusted to the MULTISENSOR Use Case Scenario to receive research results of equally high quality as for Social Sensor.

themes by finding appropriate 'case studies'. Engaging in social networks and communities they are also working as representative for their news organisation.

2.2 Specific topic: European energy policy

We have chosen "European energy policy" as a specific topic in order to illustrate how MULTISENSOR could be used in a working environment. This topic will also serve as the underlying practical use case for developing and testing the MULTISENSOR prototypes over the course of the project.

The finite nature and instability of fossil fuel supply as well as the accelerating climate change have made energy policy to one of the main issues in international news coverage. Furthermore, the long lasting struggle about nuclear energy has politicised large parts of the population in many European countries. For instance, the disaster at the nuclear power plant in Fukushima in 2011 has prompted the conservative German government to cede its support for nuclear energy. Since then, all parties in the German parliament support the "Energiewende" (energy turnaround), the fundamental policy change that includes the shutdown of all German nuclear plants as well as a massive development of renewable energies.

Despite this general consensus, the issue itself remains highly controversial. Politicians, NGOs and citizens struggle about the right strategy: Who should be burdened with the massive costs? Is there still a need for new coal-fired power stations? How can energy be stored for times when neither the sun nor the wind are providing enough output? How can electricity be transported from the windy shores in the North to the industrialised and energy-consuming South? Is "Fracking" a suitable technology? Furthermore, some European countries that are connected to the European power grid are concerned about the stability of the grid. Other countries fundamentally disagree with the shutdown of nuclear power plants and complain about Germany's unilateral decision. And the European Commission is questioning several mechanisms that have been introduced by the German government to support the "Energiewende" as illegal subsidies. On the other hand, some European countries that fundamentally support the German policy are watching this process very closely in order to adapt their national policies accordingly.

Foreseeing that this process will last several years, media organisations and journalists are interested in regular updates on European energy. They are interested in official sources such as national governments and European institutions. But they also would like to follow media outlets from different countries as well as academic institutions that deal with energy policies. Furthermore, it is of particular journalistic interest to follow comments from the civil society, including statements from NGOs and regular citizens through social networks and blogs in different languages. The goal is to give a broader perspective of different positions about energy policy in Europe and to cover the development of national concepts as well as what is going on at a European level.

Ideally, a journalist would be looking for the following information:

- A weekly list of content items from different sources relevant to energy policies in Europe.
- A summary of the most important developments in European energy policy.

- An overview of how recent developments in European energy policy have been perceived by the population.

2.3 Elicitation of user requirements

2.3.1 Methodology

The MULTISENSOR project team has followed a mixed approach of gathering the requirement for its research. The exact measures depended on the use case and the target audience. With regard to the journalistic use case, the project team has conducted a thorough market analysis of existing tools to see which other applications are currently available. The analysis has revealed best practices regarding functionality as well as UI-Design which have affected the requirements for MULTISENSOR.

Furthermore, the MULTISENSOR project team has interrogated a number of news professionals who work with large data sets and social media content on a daily basis. It was ensured that all three roles (editors, reporters and social media specialists) were represented in the survey. These interviews have led to very interesting and helpful qualitative insights. But due to a limited number of available candidates and the limited time they were prepared to spend, these results are only moderately quantifiable.

Very specifically, Deutsche Welle has:

- registered tools and processes currently in use within Deutsche Welle
- conducted a limited number of in-depth expert interviews regarding existing processes and requirements in the journalistic work structure.
- organised a focus groups with DW journalists, digital media experts and media managers to evaluate new possible requirements.

2.3.2 Competitive market analysis

As one step of the requirements analysis phase for MULTISENSOR, DW conducted a thorough market analysis of tools currently available for online research, media monitoring and analysis. In order to cover the whole spectrum of useful tools & functionalities DW relied on three channels:

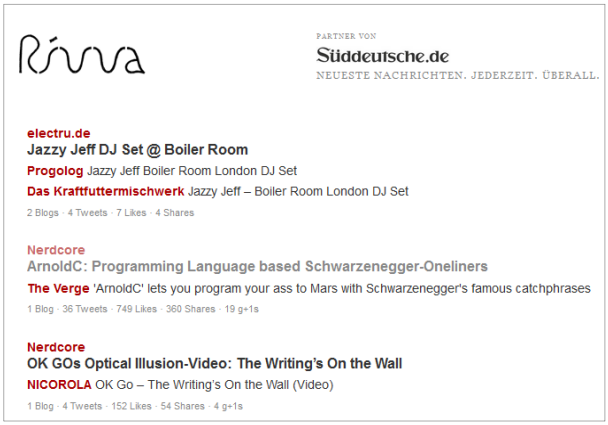
- It made use of the expertise of its own journalistic network, interviewing its members about their preferences.
- It tapped into different social media channels and
- it conducted a general online research, focusing in both cases on the main research areas of the MULTISENSOR project.

All three streams resulted in a list of different tools, covering different aspects of the envisioned MULTISENSOR solution. The tools were in most cases free or open source tools, available to everyone. In some cases a registration for a test-account was necessary. All tools were categorized according to functionality and usage area, giving a good overview of what's currently available and possible. (This complete list categorized by functionalities and including additional tools can be found in the annex of this document. Due to the constant growing number of online tools this market analysis can only highlight the functions covered by these tools. It is not considered a concluded list.)

The full list was then analysed in regards to performance, usability, usefulness for the journalistic use case and overall strength and weaknesses. By categorizing the tools into different groups, certain aspects became visible instigating a closer look at certain design elements & functionalities allowing for the definition of requirements for the MULTISENSOR development.

Social News Aggregator

In the category of social news aggregators, tools like **virato** or **rivva** help people by showing them, what's currently the most discussed topic on the web. The aggregators do so by analysing the number of shares and likes an article gets on the web via social networks like Twitter, Facebook or Google+, sometimes also including the number of mentions in blogs and on other websites (see Figure 1).



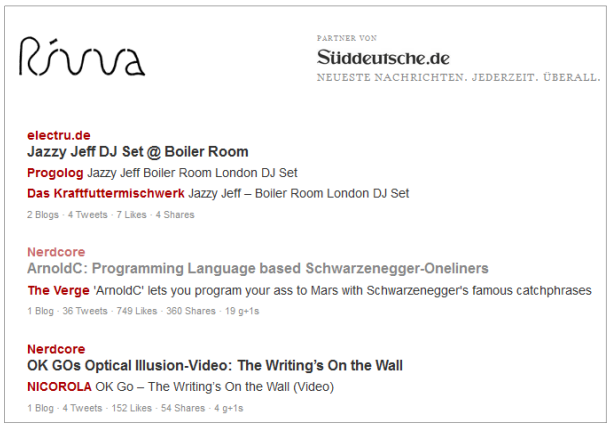


Figure 1: Social News Aggregator sites (Rivva & Virato), showing ranking, mentions and shares.

Technically interesting for MULTISENSOR is the way these sites crawl the web as well as the connected social metadata (likes and shares) collected throughout the crawling process. Furthermore MULTISENSOR can use the visualisation of the results to generate ideas on how it will structure its own search results.

Twitter Network Search and Analyses

With Twitter being one of the big and the most openly accessible social networks on the web, there are a lot of tools around for insight search and analysis. As the importance of social discussions is growing more and more important for the journalistic work, MULTISENSOR had a close look on websites like **topsy** and **twitonomy**.

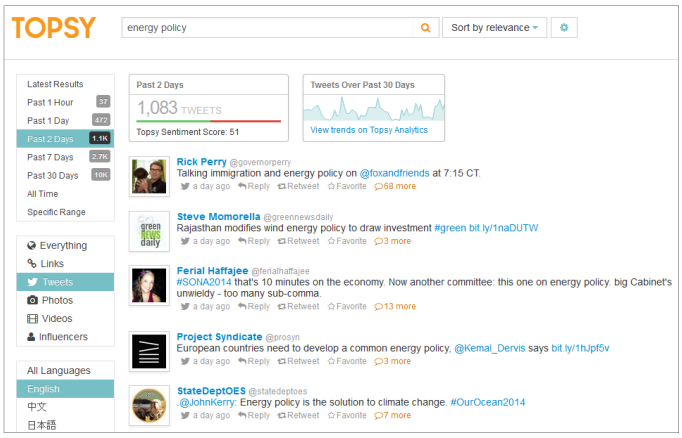


Figure 2: Topsy, with its filter options on the left

"topsy" (like other similar applications) is a social media search tool with analytic visualisations and a focus on twitter. It allows users to search for keywords, just as MULTISENSOR will do, and lists findings from twitter in multiple languages within a certain time frame. These criteria as well as some additional ones can be adjusted to tailor the result list according to what users need. This allows looking for tweets in e.g. English, from within the last week if they include a picture (see Figure 2). It even includes some graphical analysis over time as well as a very simple sentiment analysis of the findings.

Tools like *"twitonomy"* on the other hand are used to analyse twitter traffic and follower activities of a specific twitter account, helping to identify popular messages (through Retweets and Favorites) and influencers among one's followers.

Both categories serve as examples for MULTISENSOR in terms of how to build well structured interfaces as well as a good working social network analysis. This is important in order to display results in a meaningful and clear way for the user, but also to get the right information from the social network structures and find linked information. (The analysis also confirmed the necessity of different filter options for the results, also shown in the expert interviews.) MULTISENSOR will even go a step further by offering automated translation of results.

Online Media Monitoring Services

As MULTISENSOR is not only focusing on social networks but also the net in general, the analysis also encompassed tools with a stronger focus on classic media monitoring but for the online world.

There are systems like **mention** (now called alert.io) that function similar to Google alert. They allow for a keyword search across the web, alarming you about every hit they find, including social media. A more classical version are sites like **brandwatch**

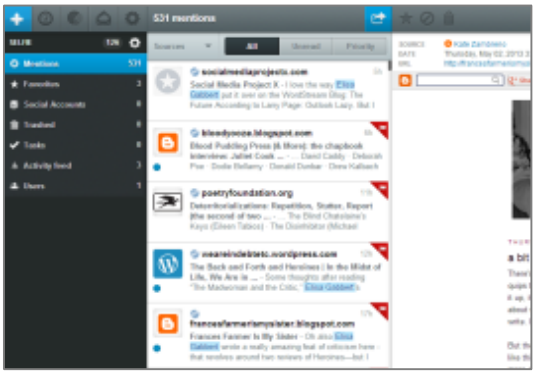


Figure 3: Listing of findings in mention

with a stronger business focus to monitor the competition. Both options are interesting due to their crawling procedures and the way they visualise their results and in depth analysis (like sentiment analysis). As there is a lot of information attached to each result, comparing different styles will be very helpful for MULTISENSOR to find the most appropriate one, without overloading its interface.

The third set category takes a closer look at traditional media sources (major newspapers like the New York Times) and offers language filters and entity extraction (tools like Newsexplorer). While the User Interface might not be the most beautiful one around, the way this application handles the news is quite a good example for MULTISENSOR. Especially the applied categorisation and presentation of metadata.

Text analysis, extraction and comparison

As the project is aiming at a deeper understanding of text, the market analysis also included tools and APIs like churnalism, opencalais and AlchemyAPI. The core requirement in this case is Natural Language Processing (NLP), comparing of texts as a whole or as fragments and extraction of entities for further analysis.

Figure 4: NewsExplorer allows for the choice of categories and entities from the content

The comparison of the application clearly showed that these processes are already being used, but rarely in combination. MULTISENSOR will go beyond this point by taking into account how these applications are handling the information extraction and how they work with the information gained. MULTISENSOR will then use this knowledge to come up with a single solution, combining these functionalities into one smart and more effective system.

Web News Filter

Another example for a possible interface is the Semantic Wire, a tool that works similar to Yahoo pipes. It lets you combine filters and output visualisations, allowing for an individual experience. For MULTISENSOR this might be another interesting example for building an interface, but also for the modular structure of its services. This is especially interesting for a journalistic use case, as every story requires its own approach.

Figure 5: The configurable interface of Semantic Wire

Conclusion from Market Analysis

The comparison of the above mentioned tools and their functionalities led to the conclusion that there is not one single tool covering all functional aspects envisioned in MULTISENSOR (see Figure 6). However the market analysis allowed for a good overview of what functionalities are currently being offered to users and what future solutions would have to comprise. The comparison of the applications also showed design aspects that MULTISENSOR will have to take into account if it wants to serve as an all-in-one multi-functional tool for journalists. This applies especially to the combination of all the different analysis elements, but also to simple elements such as the elements of a search screen.

		automatic language detection and translation	multiple source integration	analysis: sentiment	analysis: semantic	analysis: text structure	analysis: networks	entity detection	categorization	enrichement	automatic summarisation
Social News Aggregators	Virato	NO (only interface languages)	focus on 1) twitter, 2) facebook, 3) g+ and 4) Blogs	no	no	no	yes	no	yes	no	no
	rivva.de										
Network and Search analysis	twitonomy	NO	twitter only	no	no	no	yes	no	no	no	no
	topsy.com	NO (only interface & content languages)	twitter only	yes	no	no	no	no	yes (content)	no	no
Online MM Services	mention	NO (only interface languages)	twitter, facebook, google, the web in general	no	no	no	no	no	yes (manual)	no	no
	brandwatch	NO (only interface languages)	twitter, facebook, google, linkedIn, feedly	no	no	no	yes	no	no	no	no
	newsexplorer	NO (only interface & content languages)	newsites	no	no	no	no	yes	yes	offers related content in other languages	no
Text analysis, extraction & comparison and web news filter	churnalism	NO	manual input (text)	no	yes	yes	no	no	no	no	no
	opencalais	NO	manual input (text)	no	yes	yes	no	yes	yes	no	no
	alchemyapi.com	language detection but NO translation	manual input (text)	yes (entity based)	no	no	no	yes	yes	no	no
	gate.ac.uk	analysis of different languages	manual input (text)	no	yes	no	no	yes	yes	no	no
	clusty.com	NO	the web in general	no	no	no	no	yes	yes	no	no
	semantic wire	NO	the web in general	no	no	yes	yes	yes	yes	no	no
	MULTISENSOR	**WP2**	**WP 2 + WP 4**	**WP 3**	**WP 5**	**WP 3**	**WP 3**	**WP 5**	**WP 5**	**WP 5**	**WP 6**
	Journo-Interest	**high**	**high**	**low**	**medium**	**high**	**high**	**medium**	**medium**	**high**	**high**

Figure 6: Comparing planned MULTISENSOR functionality with tools currently available

2.3.3 **Interviews**

We interviewed 15 journalists and digital media experts over a period of three months with a series of in-depth interviews based on a common set of questions. The questionnaire is included as an appendix to this report. As MULTISENSOR is focussing on the analysis of mostly digital information with a considerable involvement of social media content, we have mainly selected media professionals who are already experienced in the area of online and data driven journalism. Thus, we were able to skip introductory explanations about established processes and tools that are already used by cutting edge online journalists. We also decided to follow a purely qualitative approach by interviewing a smaller number of experts but having more in-depth feedback than talking to a high number of potential users with less experience.

The questionnaire and interviews were focussed on the MULTISENSOR R&D framework as well as on testing and further elaborating specific hypotheses that could be derived from the general market expertise by MULTISENSOR user partners and particularly by DW with regard to the journalistic area. These hypotheses are built on the perception of specific challenges that journalists are obviously facing in a digital media environment: journalists fundamentally depend on the availability of information and data in order to find out what is going on in the world. On the other hand, the sheer amount of available data makes it nearly impossible for anyone to identify, process, understand, analyse and relate all this information manually and in real time. Furthermore, this information is very complex and heterogeneous; it is available in different languages and idioms and the context is often not very clear. Without having any means to identify and extract relevant information, to understand what it means and to put it into context, journalists have difficulties to benefit from available information within an appropriate time frame. This could ultimately foil the advantages of information gathering from the Internet, Social Networks and Open Data.

Considering the current situation, our hypotheses were that journalists need support in:

- understanding the specific meaning of a textual document independent of the language, tone and idioms in use.
- summarising the content of very complex and heterogeneous texts and bringing out the most important information of the original document.
- identifying related content and in combining textual data with other related data, especially multimedia items.
- finding out more about the contributors of specific information as well as about how a story (a news event, a promotional campaign) evolves over time.
- visualising social media conversations according to relevancy or other criteria of choice.

We have used the interview and market analysis process to elicit how these journalistic needs translate into very specific user requirements for a potential MULTISENSOR system.

2.3.4 General Questions

In general, participants stressed that their needs very much vary based on the topic of research. One interviewee pointed out that any journalistic research has elements of trail and error and resembles

⇨ *a hermeneutic process where first results trigger further more focussed research and so on.*

With regard to the multilingualism of potential sources, most participants underlined the need for a reliable **translation** of texts from different languages. The majority did not mind whether these texts were translated into their respective mother tongue or into English as a first common step. Many interviewees liked the possibility of **automatic summarisation** to support the research process provided that results are reliable:

⇨ *Summarisation could be helpful at the beginning of the research process: When dealing with large numbers of articles it would be great to automatically exclude doubles.*

⇨ *Summary of several texts could be useful to get a start into a topic when dealing with large amounts of data.*

⇨ *The underlying approach should be transparent so that journalists can assess quality of results.*

Sentiment analysis on the other hand, met some fundamental scepticism:

⇨ *Journalists don't really trust automated sentiment analysis because it often doesn't work very well and everybody has his own opinion anyway.*

⇨ *Sentiment analysis is overrated.*

⇨ *This is more for business reasons – not journalistic work.*

Furthermore, for some of the participants, the existing sentiment analysis tools were too simplistic:

- ⇨ *Simple sentiment analysis like positive/negative/neutral is not very interesting. But it would be very interesting to better understand certain nuances such as irony or sarcasm.*

With regard to which **additional information** MULTISENSOR should be able to **extract,** nearly all participants stressed in one way or another that this depends on the specific research topic. Nevertheless, the following pieces of information were labelled to be very interesting for most interviewees:

- ⇨ *URLs*
- ⇨ *Name of website/blog/other source*
- ⇨ *Name of author*
- ⇨ *Background information about author*
- ⇨ *Date of content creation*
- ⇨ *Geographic provenance of content*

Nearly all participants stressed the **usability** of the system as a key requirement. This mainly referred to the **configurability** of research parameters that journalists should be able to choose individually according to the topic in question. With regard to the iterative research process, participants also asked for the possibility to change or refine initial results without loosing the initial result list. Another aspect of usability where the **interface** and the ability to **visualise** data streams and correlations between different pieces of information:

- ⇨ *There should be a graphic/visual connection between summary and original text.*
- ⇨ *Extracted information and original content should be connected by using a mouse-over functionality.*

When asked specifically about the preferred visualisation, most participants mentioned **network-graphs** that show how different pieces of information and different contributors correlate.

2.3.5 The summarisation experiment

With regard to the specific attention the MULTISENSOR project as well as most interviewees have paid to a potential summarisation service, we have conducted an experiment by confronting participants with five different versions of summarisation of a text they had to read before. The five versions consisted of:

- Two phrase-extraction-based summaries
  (one created automatically and one created manually).
- Two sentence-extraction-based summaries
  (one created automatically and one created manually).
- One abstraction-based summary
  (created manually)

The original article and the different versions of summarisation are also included as an appendix to this report.

The (on first sight) surprising result was that nearly all participants dismissed the abstraction-based summarisation. Although they recognised the better readability of this summary, most of them complained about missing details:

- ⇨ *Many key factors missing in favour of readability.*
- ⇨ *Ok as an entry point, interest to read more, but numbers should be included.*
- ⇨ *In comparison to the length of the summary there is not enough information.*

One participant expressed a fundamental concern:

- ⇨ *Journalists would not use this. They would never trust an abstract summary that is created by software/an algorithm.*

With regard to the other four versions, most participants clearly preferred those that had been created manually with a slight preference for the *sentence-extraction-based summary*. The two versions that had been created automatically were criticised for different reasons. The *sentence-extraction-based summary* was perceived by some as

- ⇨ *too confusing, too detailed for a quick read, too long.*

The *phrase-extraction-based summary* on the other hand, was recognised for having the right length to give a quick overview but also criticised for appearing as a random sequence of more or less relevant excerpts from the original texts.

The findings of this first experiment will serve as a starting point for further research and development of the MULTISENSOR summarisation module. This experiment has also proven that research and development in this regard needs to follow an iterative approach where improved summaries need to be frequently validated by users in order to give feedback to technical work packages and to trigger further improvement.

2.4 Core user requirements

On the basis of market analysis and interviews we see a possible strong USP for MULTISENSOR from a journalistic point of view in

- ⇨ *an automatic summarisation of heterogeneous and multilingual digital information in English.*

This refers to text but equally so to video and audio files as they are very time-consuming and difficult to understand “manually”. Based on this summarisation, MULTISENSOR should automatically suggest related content and information that allows journalists to enrich their coverage of a specific topic. Apart from these two main functionalities (modules), the general usability of the MULTISENSOR system and its interface is highly important. This includes the possibility to individually configure and re-configure search parameters and to demonstrate results by intelligent visualisations.

More specifically - from a journalistic point of view – MULTISENSOR should meet the following user requirements:

Level 1 requirement	Description	Level 2 requirements	Priority[2]
User defined search	Search of predefined online sources (e.g. news websites, websites of public authorities and institutions, social networks, databases including audiovisual data) according to users' interests.	1. Search in full text (articles), comments (WebPages, PDFs & word) and in audiovisual data. 2. Search according to specific keywords and concepts (i.e. synonyms and hyponyms should be automatically considered). 3. Search in specific media sources or media types. 4. Search in specific languages 5. Search according to location 6. Search within temporal dimensions. 7. The system must provide a list of the results, each of them represented by a small summary. 8. Item similarity search: the user should be able to select an article/item and retrieve relevant items.	1
Automatic language detection and translation	Software-based language detection and translation of text documents (main focus) and speech from multimedia from one language to another.	1. Identification of language in articles and multimedia. 2. Translation of textual and speech content to language of choice	1
Automatic summarisation	The software-based process of reducing a text document or other content in order to create a summary that contains the most important information of the original document.	1. Summarisation of long texts. 2. Summarisation of text bundles. 3. Summarisation of audio-visual content. 4. Summarisation of social media threads and conversations (*priority 3*).	1

[2] meaning 1 = highest priority and 3 = lowest priority

Level 1 requirement	Description	Level 2 requirements	Priority[2]
Sentiment analysis	The software-based process of identifying the sentimentality (aka subjectivity) and the polarity of a piece of text, which reflects the author's opinion towards a specific subject. Sentiment analysis can be performed on a document-, paragraph-, or sentence-level, and can be specific to the entities, concepts, brands, or products mentioned in the text. The output is a set of scores in a predetermined scale.	Sentiment analysis with regard to different entities that are mentioned in one text.	3
Named Entities Extraction	The software-based process of extracting named entities from text and audio for further analysis.	Extraction of 1. URLs 2. Name of website/blog/other source 3. Name of author 4. Names, brands, geo-locations, etc mentioned in the textual and audio source.	1
Context extraction	The software-based process of identifying background information and relating it to other pieces of information.	1. Analysing background information about • author • date of content creation • geographic provenance of content 2. Identifying content and information that is related to specific extracted entities.	1
Enrichment	The software-based process of combining (mainly textual) data with other related data, especially multimedia items identifying common characteristics between different items (e.g. topics, events) and identification of hidden meanings and correlations based on the detection of complementary, contradictory and similar content.	1. Identification of related social media content. 2. Automated clustering considering different dimensions (e.g. semantics, spatiotemporal) 3. Suggestion of similar pieces of information from different sources (especially videos, images and multimedia items). 4. Detection of facts that are contradictory/complementary or similar mainly in different items.	1

Level 1 requirement	Description	Level 2 requirements	Priority[2]
Contributor analysis	The software-based process of collecting information about the contributors of information and data (e.g. name, origin, background, history, influence scores, position in network).	1. Detection of personal information about specific contributors (age group, nationality etc.). 2. Analysis and display/ visualisation of contributor networks. 3. Calculate influencer scores. 4. Database of identified relevant contributors.	2
Configurability of system results	The user should have several options to display MULTISENSOR results according to specific needs.	1. Filtering by entity (concept and Name entity) 2. Reducing to fewer entities. 3. Saving lists 4. Combination with other lists.	1
GUI development	The MULTISENSOR interface should be simple to use and allow individual configuration.	See mock-ups in this deliverable.	1
Visualisations	Exploring and implementing ways of visualising MULTISENSOR analysis results in an understandable and user-friendly way	1. Visual connection between summary and original content. 2. Visual or other (e.g. mouse-over) connection between extracted entity and original content. 3. Network graphs of contributors and/or different content items.	2

Table 1: Requirements overview Journalistic Use Case

2.5 Summary and priorities

Challenges for and needs of journalists are manifold when it comes to dealing with large amounts of digital data. Given the limited scope and timeframe of the MULTISENSOR project, each use case scenario should focus on the most important requirements. From the journalistic point of view, this is

⇨ *automatic summarisation of heterogeneous and multilingual digital content (text and audio/video) in English.*

If possible, MULTISENSOR should also automatically provide journalists with content and information that is related to the topic of choice. Apart from these technical requirements, usability and configurability of the system are highly important as well.

3 PILOT USE CASE 1 (SCENARIO 2): COMMERCIAL MEDIA MONITORING

3.1 Targeted user groups and usage context

The second scenario of PUC1 targets media monitoring companies and their clients. The market for commercial media monitoring has experienced huge changes in the last decades with the advent of - first - online media and the more recent rise of social media and the huge amount of data it produces. Coming from a background where media monitoring companies used to cut clippings from printed media and mail them to their clients, the competitors nowadays need to completely cover and analyse multiple media sources in near real-time in order to succeed on the market. Especially the large amount of social media content calls for at least partly automated processes. The project team has identified the following relevant personas in the commercial media monitoring context:

- **Media monitoring company, editor**:
  In their daily work, editors working at media monitoring companies are confronted with the challenge of having to filter and analyse huge amounts of data. The framework for content selection is predefined by the customer's need. Depending on the covered sector a lot of special market knowledge is necessary. To help them accomplish their tasks, editors are interested in system support that can reduce the number of decisions they have to make e.g. in article selection. For them, it is important that all data is presented in an intelligent and structured way that allows the editors to grasp the most important information at little more than a glance.

- **Media monitoring company, analyst:**
  Analysts in the media monitoring business are responsible for extracting the most relevant information from the coverage a client expresses interest in. They are charged with assessing the sentiment, topics and key influencers present in the discussions around their clients. Thus, they need to accurately handle large amounts of data and extract the most interesting turns and twists in the coverage. An analyst's work is about producing insight for the client, and thus these persons are highly interested in intelligent visualisation techniques.

- **Media monitoring client, management level:**
  These are executives at large, possibly diversified and/or international companies with a lot of responsibility and little time available. From their media monitoring provider, this group expects condensed information that informs about the main trends of the company's media coverage. These persons are also interested in benchmarks that give them a quick overview of the competitive environment. To them, an individual piece of news is of less interest than the overall achievement of strategic communication goals, which might be the dissemination of key messages and a wide dissemination.

- **Media monitoring client, project level:**

PR managers at client companies are often in charge of specific campaigns or product segments and wish to measure their own PR work through the media monitoring and analysis services they book. To them, a wide and positive dissemination of news and messages is crucial, so they emphasize the importance of the completeness of the coverage. The crawlers need to be reliable and results visualised in an attractive way, as this is how the project managers present their work to their superiors.

3.2 Specific topic: Media coverage of the household appliances market

The topic in this scenario depicts the media monitoring and analysis needs of a large diversified household appliances manufacturer, who signs up as a client with a media monitoring company.

Household appliances are everyday items that virtually everybody uses. They facilitate people's lives and are prominent parts of their homes, especially in kitchens and bathrooms. This wide distribution also means that most people have an opinion on the products, which leads to an enormous amount of user generated content on the social web as well as many mentions in the popular press. Important topics for household appliances are for example functionality, design, environment friendliness and energy consumption.

In the market, there is a lot of competition, and the market leading players need to have consistent strategies regarding innovation and cost issues. Like all companies nowadays, they need to concern themselves with social responsibility topics, as these may directly influence consumer behaviour.

For our use case scenario, we assume that a leading manufacturer of household appliances wants to use the MULTISENSOR platform in order to access and organise its entire global media monitoring and analysis activities, both in traditional and in social media.

The client requests that all data must be mapped according to country and business sector (i.e. major appliances, floor care, small appliances etc.). At the same time, a portal needs to be set up that calculates key performance indicators from the media monitoring results of the different regions and sectors.

The client requests the MULTISENSOR platform to map all articles not only to business sectors, but also to a multitude of specific products and marketing campaigns that accompany the launch of new products. It is further agreed to work with a given set of qualitative analysis parameters such as target media, tonality, message penetration, picture penetration, exclusivity and mentions of corporate spokespersons. In addition the client is interested in identifying media contacts to reuse these in a contact database.

For the sake of benchmark analysis articles from target media are to be examined to identify competitor brands. The client also wants to track consumer opinions and sentiment with regards to specific products, as well as predict market trends and customer behaviour based on similar situations in other countries.

Confronted with these tasks and a huge amount of coverage, a media monitoring company needs to set up a workflow with a streamlined selection process for relevant data. This would ideally include

- a search engine that presents all relevant content in an intelligently clustered way from which a selection can be performed.
- an on-demand translation tool that helps already during the selection process.
- a summarisation tool that can compress the content of an article or a social media discussion and identify the most important passages.
- an organisation tool that clusters the selected data automatically or semi-automatically according to the client's needs and is able to consider these (self-created) categories or codes as indicators for analysis.
- a contextualisation tool that on-demand provides additional content like article commentaries.
- a contributor analysis that can evaluate an author's impact in the network and provide background information on him.
- several visualisation techniques that can help with the analysis ordered by the client and that show the connection of topics and news.

3.3 Elicitation of user requirements

3.3.1 Methodology

As mentioned before, a mixed approach was chosen for elicitation of the requirements for MULTISENSOR. For the media monitoring use case, the project team performed an in depth market analysis consisting of three pillars:

In a first step, we reviewed currently running media analysis projects at pressrelations. These real-life client projects helped to assess what features are of most importance to companies of various sizes and with different budget limitations.

The second step consisted of a competitive market analysis of the media monitoring market. The project team reviewed the member list of the biggest associations in the sector, namely AMEC[3] and FIBEP[4]. Here, we focused on technology providers for media monitoring companies, as this is one of the market segments where MULTISENSOR will most likely position itself. Special attention was given to innovative technological solutions. Additionally, we considered the market-leading media monitoring companies such as Meltwater, who have developed their own SaaS solutions. This approach has revealed similar technologies already being used in the media monitoring market as well as possible gaps.

Thirdly, we conducted a small number of structured in-depth interviews. As interviewees, we went for key account clients and partners of pressrelations, who are of special importance because they could bring in the experience with their own client pool.

These are the measures pressrelations performed in an overview:

- We conducted a literature research for the media monitoring sector.
- We analysed our internal work processes.
- We assessed our current clients' analysis projects.

[3] International Association for Measurement and Evaluation of Communication; http://amecorg.com/.
[4] Fédération Internationale des Bureaux d'Extraits de Presse; http://www.fibep.info/.

- We performed a market analysis of system providers for media monitoring companies.
- We conducted a limited number of in-depth interviews with key account clients and partners in the media monitoring industry.

3.3.2 Market analysis

Multilingual Technologies

In the category "multilingual technologies", SAIL LABS, Meltwater and TVEyes offer various solutions to problems that are dealt with in MULTISENSOR.

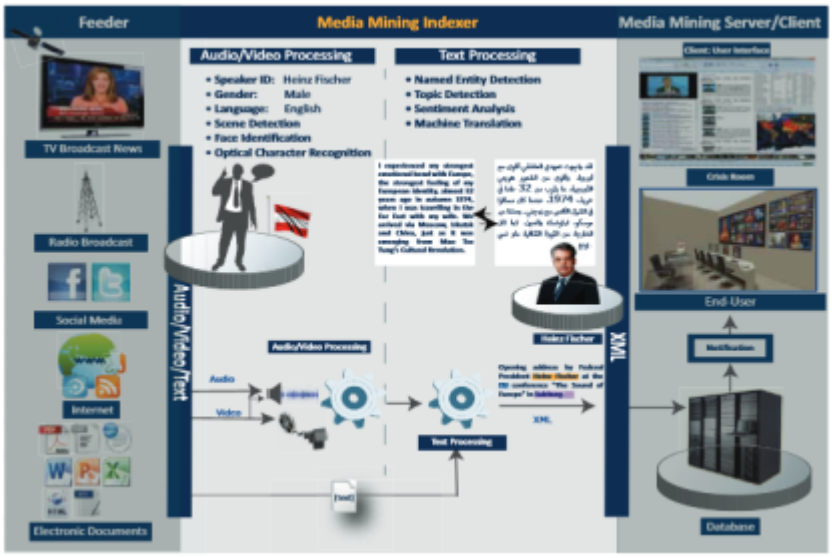


Figure 7: Screenshot Media Mining Indexer

In the field of speech technology, **SAIL LABS** ("Speech-Artificial-Intelligence-Language-Laboratories") is one of the world's leading innovators, creating high-end software for speech and multimedia analysis solutions. The Media Mining Indexer is based upon SAIL LABS' own speech recognition technology and provides a suite of best-of-breed components for multimedia processing, transforming audio and video data into indexed and searchable information in real-time.

SAIL LABS' real-time multilingual automatic speech recognition technology (Automatic Speech Recognition (ASR)) offers the following functionalities:

- Keyword Translation, supported languages: Arabic, Bahasa Indonesia, English (US/UK), Farsi, French, German, Greek, Hebrew, Italian, Mandarin Chinese, Norwegian, Polish, Romanian, Russian, Spanish
- Automatic Speaker ID / Clustering: The Speaker Identification (SID) system identifies speakers or the speakers' gender
- Machine Translation; Language Model Toolkit (LMT): build own vocabulary
- Data can be automatically indexed and enriched e.g. by Named Entities Detection, Topic Classification and Categorisation, Linguistic origin of speaker for TEXT MINING INDEXER Data Cleaning (the relevant portions of a text document are extracted and the text is standardized e.g. expanding abbreviations), Story Segmentation (the text output is

segmented into coherent stories), Topic Detection and Sentiment Analysis (positive, neutral, negative)

SAIL LABS offers products relevant to MULTISENSOR in the categories speech-to-text-conversion, automatic multilingual transcription and machine translation, speaker identification and automatic data analysis.

In competition with SAIL LABS we find **Meltwater's software** which attempts to understand the message and sentiment of articles from 17 different languages.

TVEyes offers real-time transcription and translation from multiple languages into English and the ability to easily translate between multiple languages. The supported languages are UK English, US English, German, Spanish, Italian, French, Chinese, Greek, Turkish Russian and Arabic.

Sentiment and context analysis

In the category "sentiment and context analysis", there are two major software solutions available, offered by iQ Media Group and again, Meltwater.

iQ Media Group offers an algorithm for detecting tonality in a document, using the words surrounding the search term. The algorithm is based on the quality of language surrounding the search term and the determination of the tonality of the words that are in proximity to the result. Rather than allocating a three-bucket approach (positive, neutral, negative) a score approach was instantiated. The scores can be customized for clients based on their needs. The score allows us to differentiate the determination of positive or neutral, negative or neutral. Some clients call positive sentiment as anything that is not negative and vice versa. Score cut offs can be based on the client's needs. Also, client specific terms or phrases can be picked up that are specific to their industries.

Figure 8: Screenshot iQ Media Group

Meltwater provides the online intelligence solutions Meltwater NEWS (Online) and Meltwater Buzz (Social Media) as SaaS solutions with the claim "Just set it and forget it". Meltwater offers sentiment analysis, a summary of articles by positive, negative or neutral tone, extraction of topical themes, volume and geographic data. In detail, the following features are available:

- Volume – Monitor reach of social conversations about companies, brands, campaigns, competitors, and industry.
- Media Spread – Track the location and spread of the conversation on Twitter, Facebook, blogs, forums, comments, and media sharing sites.
- Sentiment – Understand the tone of the conversation as positive, negative or neutral.
- Themes Cloud – Analyse the underlying topics of conversation and drill down to the individual messages to find new conversations to monitor, content to share and campaign ideas.
- Word Cloud – Visualisation of topical themes in recent press.
- World Map – View a conversation's distribution by country

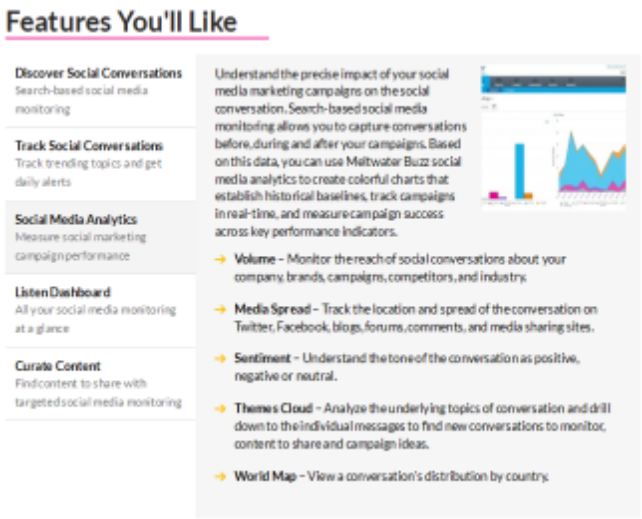


Figure 9: Screenshot Meltwater

One competitor of Meltwater is Business Wire Inc. which offers the distribution of press releases and per-press-release reporting (ONLY for English-language press releases) on social media volume, sentiment, major influencers, geographic location and other data.

Crawling infrastructure

In the category "crawling infrastructure", NEWBASE and again Meltwater offer solutions for media monitoring purposes.

NEWBASE also offers a solution for crawling data, namely the NEWBASE iSpider. Among the functionalities of the crawler are

- Fully automatic observation of defined websites. The program analyses websites at pre-set intervals and identifies new articles.

- Fully automatic capture of WebPages. If a new article is discovered, the program saves the page including metadata (name of the website, topic, author etc.).
- Fully automatic article clipping. In the enhanced version of the program the article is clipped from the webpage whereby all irrelevant elements such as advertising, menus etc. are removed automatically. Metadata (name of the website, topic, author etc.) are also saved.

Figure 10: Screenshot NEWBASE

- Fully automatic saving and classification. The WebPages or articles are saved to the database. If you maintain reader profiles, the program automatically searches for key words in the contents. The articles are allocated to the topics on the basis of the key words found.
- Release in the portal. The Internet articles can be presented in the NEWBASE Media Portal along with other articles from various media channels. In other words, all articles are available in one Intranet.

iSpider is relevant in the context of MULTISENSOR since it attempts fully automatic media monitoring (complete solution provider for media monitoring software, digitisation and conversion services and app development).

Meltwater also offers solutions in the area of data crawling. The Meltwater Buzz Listen module's search-based social media monitoring tool uses keywords to identify specific topics of conversation that are important to the community, such as brand names, competitors, industry phrases, acronyms, social hashtags, celebrities, and trending topics, using precise Boolean logic.

Meltwater News online media monitoring facilitates the same type of search for standard online media. Its crawling solution is relevant in the context of MULTISENSOR since it aims at finding the most relevant content in social and online media, using automated, keyword-driven search agents.

Graphic interfaces

Graphic interfaces for media monitoring and analysis are offered within the software solution by **iQ Media** which is a Media Intelligence Platform, enabling users to leverage TV video, online news, and social media content optimized for the cloud. Users do not need to invest in IT infrastructure or management of a network, but they have instant access to media at all times from any Internet-capable platform.

Figure 11: IQ Media Dashboard Interface

The software solution cliQ 4.0 has been developed with the notion that big data can be ingested, organised, and made available. cliQ's cloud-based media is based on a software developed by iQ media, the Lakshmi Scalability Algorithm (LSA).

The following functionalities are covered by iQ Media's software solution:

- cliQ: Cloud based: Media is optimised for availability, accessibility, and quality. cliQ makes media available far longer, accessible far more quickly, and presented in a higher quality format.
- Total Recall: Media is available for longer periods — even indefinitely. Dated TV coverage and online news that suddenly become important can still be accessed months later in the cloud.
- Fast Federated Searches: Content is pre-optimised for rapid access across all media types. Users get results quickly, pinpoint what they need, and repurpose media for sharing and permanent storage.
- Simple Sharing: Users can quickly email links to video clips and other media, post them to social media platforms, or stream them from their own websites.
- Extended Management: Users can put their own user-generated video content into the cloud, where cliQ will apply the LSA to optimize it.

iQ Media's graphic interfaces are relevant for MULTISENSOR in the sense that they serve as a blueprint for platform setup

Summary

To sum up, it has been identified that there is no single media monitoring software to date that can cover all the areas that will be covered by the MULTISENSOR tool. Products like the ones by Meltwater, NEWBASE or the iQ Media Group cover a variety of areas that will be dealt with within MULTISENSOR, but still there is no product that combines all the functionalities that are essential for media monitoring companies and clients (as can be seen in **Figure 12**).

		Multilanguage Support	Multiple Source Integration	Analysis				entity detection	categorization	Enrichment	Automatic Summarization
				sentiment	semantic	text structure	networks				
Multilingual technologies	SAIL LABS Media Miner Indexer	keyword translation , MT	print, audio and video data	yes	no	yes	no	yes	yes	yes	no
	Meltwater	yes	print and audio data	yes	yes	yes	no	yes	yes	yes	no
	TVEyes	yes	audio and video data	no	no	no	no	no	no	no	no
Sentiment and context analysis	iQ Media	no	online and social media	yes	no	no	yes	yes	no	no	no
	Meltwater	yes	print and audio data	yes	yes	yes	no	yes	yes	yes	no
	Business Wire Inc.	yes	social media	yes	no	no	yes	yes	no	no	
Summarisation information technologies	NEWBASE	yes	social media, print	no	no	no	no	no	yes	no	yes
Crawling infrastructure	NEWBASE iSpider	yes	social media, print	no	no	no	no	no	no	no	no
Graphic interfaces	iQ Media	no	print and audio data	yes	no	no	yes	yes	no	no	no
	Meltwater	yes	print and audio data	yes	yes	yes	no	yes	yes	yes	no
	MULTISENSOR	WP2	WP 2 + WP 4	WP 3	WP 5	WP 3	WP 3	WP 5	WP 5	WP 5	WP 6

Figure 12: Comparing planned MULTISENSOR functionality with tools currently available

3.3.3 Interviews

Wishing to include multiple points of view, pressrelations conducted in-depth interviews within its partner network for the elicitation of the user requirements. As interviewees, we chose clients and partners that can be viewed as multipliers because of their own large customer base. All of the interview partners have extensive professional experience in the field of media monitoring and analysis. The interviews, four in total, were conducted either in personal meetings or via telephone.

After an initial introduction of MULTISENSOR, its aims and approaches, we discussed the existing workflows and tried to determine possible requirements and attractive features for the MULTISENSOR system. The interviews were conducted in an open manner and structured by – but not limited to – the following questions:

- As a company that is involved in media monitoring, what are the requests you are mostly confronted with?
- What are the most time-consuming factors in your daily work?
- What metrics are of special importance to your clients?
- In your opinion, what are the greatest challenges in handling social media?
- What kind of content visualisation would be of help when performing your daily work?
- What kind of product would be truly innovative from your point of view?

During the interviews, notes were taken and a protocol was written afterwards. These protocols can be found in the appendix.

The interests of the interviewees were clearly divided into the field of monitoring on the one hand and analysis on the other hand. When involved with manual or semi-manual monitoring and article selection tasks, the interview partners expressed special interest in

- ⇨ automated clustering and good visualisation of clusters with identical articles,

because this provides a better overview and minimizes reading time. The selection process would also be aided by a display of detected entities and an intelligent

- ⇨ visualisation of basic article topics, e.g. through tag clouds.

One interviewee said that instead, a short written synopsis might be useful already during the article validation process in order to get a quick grasp of article contents.

Due to the large amount of data that needs to be processed for analysis (especially when it comes to social media), it was emphasized that well-working automated processes would be hugely valuable especially for entity and topic detection, topic tracking and sentiment analysis. While fully automated coding processes usually yield less satisfactory results, MULTISENSOR could provide substantial support for human coders and analysts with its analysis results.

When preparing analysis reports for clients, growth in media activity is always highly relevant and might require comments. Thus, an accurate

- ⇨ depiction of the media activity of authors and users
- ⇨ and topic tracking

were requirements expressed by one interview partner. This is closely linked to another requirement that is especially important in the field of social media, namely a well-working

- ⇨ detection and visualisation of relevant influencers, networks and communities.

For clients, the information “who is talking about the company and its products” is crucial and any available background information on authors and (in the case of social media) users is valuable.

From an analyst’s point of view, the interview partners further expressed a need for system support when it comes to an in-depth understanding of the topics to be analysed. An idea was proposed for manually extracting topics for further analysis through tag clouds that can be navigated. Having filtered all relevant content for a given topic, the interviewees wished for

- ⇨ topic summaries as support for analysis.

Last but not least, for an easier generation of input-output-analysis that measure the impact of the client’s PR work on media coverage, the interview partners suggested a tool that enables

- ⇨ tracking of quotations and plagiarism.

The idea was to implement a function where e.g. the texts of client press releases can be uploaded and made available for a cross-check with news articles. As large text bodies need to be compared over a long time frame, this is perceived as an especially difficult task to perform manually without system help. As this, however, is no task foreseen by the MULTISENSOR DoW, we would like to introduce the tracking of quotations and plagiarism as a “nice-to-have” feature with the requirements.

3.4 Core user requirements

From a media monitoring point of view, it is important that the MULTISENSOR system follows the usual workflow for the creation of a media analysis. In a first step, the user needs to define the sources and time frame that is to be monitored along with the search terms he wants to use. This might be a one-time search or a continuous monitoring. In a second step, the search results need to be curated and validated. This step is followed by annotating the texts, e.g. by assigning codes and/or an automated analysis. Once the data is correct and complete, queries can be made. The MULTISENSOR system should in the best case present the results of these queries in different output formats and visualisations. From this background, in our opinion MULTISENSOR should address the following requirements:

Level 1 requirement	Description	Level 2 requirements	Priority[5]
User defined search	Search of predefined online sources according to user interests.	1. Continuous search profiles and one-time searches. 2. Search according to specific keywords. 3. Search in specific media sources and media types. 4. Search in specific languages. 5. Search according to location. 6. Search within temporal dimensions. 7. Search by image example. 8. Search by topic.	1
Semantic search	The software-based process of searching for discrete semantic entities rather than keywords, based on an expandable ontology. Items that are not part of the knowledge base must allow for string-based search.	1. Keyword disambiguation. 2. Upload functionality for domain specific ontology information.	2
Topic detection and clustering	The software-based process of identifying the specific topics or topic clusters of a text document independent of the language in use.	1. Topical clustering. 2. Structured display of search results.	2
User-based validation of relevance	The user needs the possibility to select relevant items from the search results while discarding irrelevant ones. Only validated news items will be subject to further processing and analysis.	1. Possibility to mark specific content items as relevant or irrelevant.	2
Named Entities Extraction	The software-based process of extracting named entities from a text for further analysis.	Extraction of 1. Persons 2. Corporations 3. Brands 4. Locations 5. Products	1

[5] Meaning 1 = highest priority and 3 = lowest priority

Level 1 requirement	Description	Level 2 requirements	Priority[5]
Sentiment analysis	The software-based process of identifying the sentimentality (aka subjectivity) and the polarity of a piece of text, which reflects the author's opinion towards a specific subject. Sentiment analysis can be performed on a document-, paragraph-, or sentence-level, and can be specific to the entities, concepts, brands, or products mentioned in the text. The output is a set of scores in a predetermined scale.	1. Sentiment analysis with regard to different entities (e.g. brands) that are mentioned on a scale.	2
Contributor analysis	The software-based process of collecting information about the contributors of information and data (e.g. name, origin, background, history, influence scores, position in network).	1. Creation of a database of authors/ contributors. 2. Detection of personal information about specific contributors (age group, nationality etc.). 3. Calculation of influencer scores. 4. Depiction of growing media activity.	1
Automatic summarisation	The software-based process of reducing a text document in order to create a summary that contains the most important information of the original document.	1. Summarisation of longer texts. 2. Summarisation of multiple texts.	1
Automatic language detection and translation	Software-based detection of source language and translation of source texts into target language.	1. Automatic detection of source language. 2. Automatic translation of documents from source to target language.	1
Further analysis	The software-based process of identifying and/or breaking down the structure, content and background of a text document.	1. Analysis by metadata. 2. Analysis by predefined indicators.	2
GUI development	The MULTISENSOR interface should be simple to use and allow individual configuration.	See mock-ups in this deliverable.	1

Level 1 requirement	Description	Level 2 requirements	Priority[5]
Visualisations	Visualisation of MULTISENSOR analysis results in an understandable and user-friendly way.	1. Network graphs of contributors and/or different content items. 2. Clickable tag clouds. 3. Pie charts, bar charts and line charts for different queries.	2

Table 2: Overview of basic requirements for media monitoring use case

3.5 Summary and priorities

From a media monitoring point of view, any system support that can speed up the reading and analysis process is valuable.

While valid and widely-used approaches exist for the monitoring and analysis of traditional media, the sheer mass of social media poses comparably new challenges to the monitoring industry. Any MULTISENSOR component should therefore be tested for its applicability with large amounts of social media content.

Apart from this general inclination, we see a big potential in the **summarisation of single articles and article bundles** that should be made possible on demand for the user at several points in the MULTISENSOR system. A second focus should be laid on **entity extraction** and **contributor analysis**.

Last but not least, **intelligent visualisations** are of the greatest help for gaining quick deep insights into analysis topics and should receive a high priority also within MULTISENSOR.

4 PILOT USE CASE 2: SME INTERNATIONALISATION

4.1 Targeted user groups and usage context

PUC2 deals with SME internationalisation which refers to the process of expanding from a regional or a national market to a new and foreign market in order to increase turnover and profit. This process is of particular importance as it is often the only option to achieve growth. But it is also aligned with considerable challenges such as a lack in knowledge about market conditions or the spoken language in the targeted countries. The target group of PUC2 are SMEs (Small and Medium sized Enterprises) based in Spain that want to start internationalising themselves. The reasons behind this decision are increasing their sales and diversifying the risk, in order to maximize their productivity and expand their market abroad. Most of these companies don't have any knowledge on how to do so or which steps to follow.

We differentiate 2 targets, which have guided our requirements:

- SMEs (Small and Medium Enterprises): while they might not directly study the market, they will have the final decision and it is important to have them in mind throughout the whole process.
- Export Managers: they are the ones that directly manage the process of internationalisation of a company. On their daily basis they are looking for the information and helping SMEs export their products, so they have a broad knowledge about opening new markets.

4.2 Specific topic: The expansion of a Spanish yoghurt producer to a new market

With regard to PUC2, we have chosen the following practical use case scenario: The Company X is a family business located in Catalonia (Spain) that elaborates artisan dairy products like yoghurt, cheese and *mató* (a sweet, unfermented fresh cheese typical from Catalonia).

They started exporting in 2010 and they are currently exporting less than 5% of their turnover. The CEO wants to increase the sales of their dairy products abroad but X has 20 employees and none of them speak properly any foreign language.

X's CEO knows that some competition is selling to the German market and he wants to know if he could sell dairy products there. To do so, he first needs to investigate whether Germany is an interesting country or not. For that, he needs a quick overview on the country's GDP, the income, the main sector and other market criteria. Since the employees do not speak German, they look up for information about the German market in Spanish and English but they have no access to the data that is only in German. Furthermore, they usually find long articles from which, after being examined, they can only use a small part.

The CEO is also interested in finding information about what people like and dislike, their opinions and social considerations, in order to know whether their product will be accepted. This data is found in different social networks.

Once they decide that Germany is an interesting country for them, they need to search regulatory issues for exporting their products there, e.g. the general and the specific legislation about dairy products. They also need to know which information has to appear on the labels, for instance. The problem, again, is that they might not be able to access that information or they might not understand it.

If they are able to find all the mentioned information, they might want to check if there is any other country that is also interesting for them. If they wanted to compare it with France, for example, they would have to do the research of a new market in an unknown language all over again, and aggregate and store all the information in excel files to be able to compare it.

4.3 Elicitation of user requirements

4.3.1 Methodology

Regarding the internationalisation use case, PIMEC has conducted a benchmark analysis of existing tools that try to solve similar problems to the ones that the MULTISENSOR platform aims to solve.

We approached several SMEs and internationalisation experts to find out what information they would need, such as a description of the country itself (population, GDP structure, social indicators…), consumer behaviour and best ways to enter the market.

We conducted two focus groups where we extracted other requirements and could see the priority of the requirements already stated.

In general, PIMEC has:

- studied which are the important matters when internationalising a company.
- conducted a research to understand which criteria need to be studied and which is more important.
- conducted several telephone surveys in order to identify specific problems that SMEs have and which information they need.
- organised a focus group with SMEs (small and medium enterprises) to evaluate new possible requirements.
- organised a focus group with export managers to evaluate new possible requirements
- held review meetings with experts and user group members to verify these requirements

4.3.2 Market Analysis

Our market analysis has confirmed that there are only a few portals and tools that are helpful for SMEs that intend to expand in new markets. National institutions such as the **Spanish ministry of commerce** or international/global institutions like **the World Bank** only offer very preliminary and basic information about the targeted country.

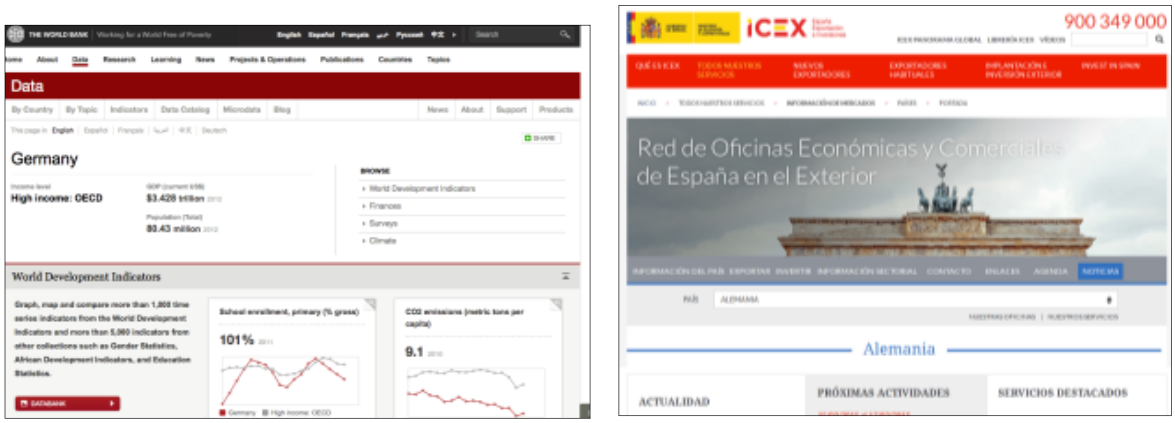


Figure 13: Websites of the Spanish Ministry of Commerce and the Worldbank

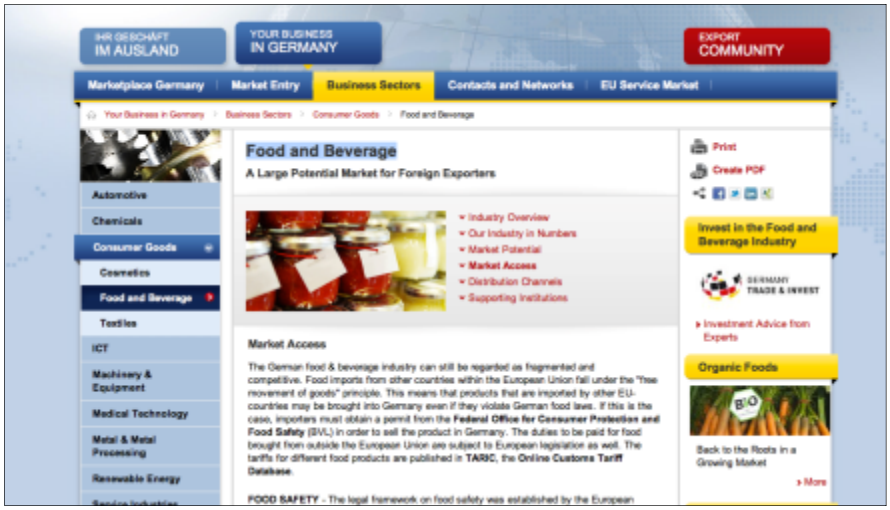


Figure 14: German Business Portal Ixpos

It is also possible to find some basic and sometimes sector-related information in English that is provided by the targeted country itself, like through the **German Business Portal Ixpos**.

But when digging deeper and looking for the crucial bits of information it is impossible to find anything that is available in Spanish or at least in English. This refers up-to-date market data, consumer statistics as well as to very specific pieces of information like relevant legislation or restrictions. The few available websites are attached in the Appendix (0). Looking at the comparison of these few tools with the planned functionalities in MULTISENSOR (see Figure 15), but also the lack of more tools clearly shows the relevance of the MULTISENSOR project.

	Multilanguage Support	Multiple Source Integration	Analysis				entity detection	categorization	Enrichment	Automatic Summarization
			sentiment	semantic	text structure	networks				
icex	no	text, videos, pictures	yes	yes	yes	no	yes	yes	yes	no
cia.gov	yes	text, videos, pictures	no	yes	yes	yes	yes	yes	yes	yes
auswaertiges-amt	yes	text, videos, pictures	no	yes	yes	yes	yes	yes	yes	yes
accio	no	text, videos, pictures, soc. media	yes	yes	yes	yes	yes	yes	yes	yes
europa.eu	yes	data, text, pictures	no	yes	yes	ye	yes	yes	yes	yes
Ixpos	no	text, videos, pictures, soc. media	no	yes	yes	yes	yes	yes	yes	yes
doing business	yes	text, videos, pictures, soc. media	yes	yes	yes	yes	yes	yes	yes	yes
worldbank	yes	text,, data, videos, pictures, soc. media	yes	yes	yes	yes	yes	yes	yes	yes
coface	no	text, videos, pictures, soc. media	yes	yes	yes	yes	yes	yes	yes	yes
comtrade.un.org	no	data	no	no	no	no	yes	yes	yes	yes
nferias	no	text	no	no	no	no	no	yes	yes	yes
eventseye	yes	text	no	no	no	no	no	yes	yes	no
MULTISENSOR	WP2	WP 2 + WP 4	WP 3	WP 5	WP 3	WP 3	WP 5	WP 5	WP 5	WP 6

Figure 15: Mapping of functionalities for Internationalisation Use Case

4.3.3 **Working committee and interviews**

The MULTISENSOR platform aims, among other things, to help the SMEs before and during the internationalisation process. Therefore, the people beneficiating from it will be both, SMEs that have some knowledge on how exports work and experts on internationalisation that work as export managers.

We believe that the information and insights that these potential users can give us is very important, so that we introduced them to the project and collected information from them.

The interviews and the "User Day" were focused on the MULTISENSOR requirements as well as on receiving information on the companies and the export managers needs when they want to internationalise a company.

We first explained some of the requirements already requested by PIMEC:

- Specific research about the market, sector and product.
- Automatic language detection and translation.
- Summarisation of information.
- Identification of the sentiments or feelings about the market, sector or product.
- Selection on which information is the most important, possibility to add it to favourites and organize them in folders.
- Comparison of the criteria depending on the country.

On one hand, we conducted interviews with representatives of different SMEs that are already exporting successfully and that have an export department.

From our conversations with the SME representatives we expected to learn about their process when researching a new market, about which information they find most important,

which sites they used, how they want to receive the information, and the most important issue: whether the requirements that we are considering would be useful for them or not.

From these conversations we got some new sites that might be important for MULTISENSOR (added under the previous section). We also understood that during the **market analysis** that they conduct before deciding which country they will start exporting to, MULTISENSOR will be of great help on reducing their amount of work with regard to the amount of barriers SME internationalisation has to face in a new market:

⇨ *Information about **market, sector and product barriers** is a must when we are looking for a new market, in order to know if it's feasible to enter or not. MULTISENSOR can provide us with a **quick overview** on whether entering a new market is hard or easy.*

We already know automatic translation which is linked to the summarisation and extraction of information is highly relevant for SMEs.

⇨ ***Summarisation** could be helpful in order to know which articles are more suitable and which are not, since we usually lose many hours going through extended articles that are not useful for us.*

Regarding which additional information MULTISENSOR should be able to extract, nearly all SMEs mentioned that it is very time consuming to **research for new clients and distributors**. It would be very useful for them if MULTISENSOR could provide information about these aspects (including their location).

When asked specifically about the preferred visualisation, nearly all participants mentioned that the **usability and the visualisation** are important for them and also that they want to be able to see it graphically.

On the other hand, in order to get insights from the internationalisation experts, we invited them to a user day at the PIMEC offices. From more than 10 participants, divided into two different focus groups, we wanted to extract some new requirements that we could add, and especially we wanted to receive feedback from experts that could be using the MULTISENSOR platform in the future.

The experts gave us some very interesting information. They see a very big potential in the MULTISENSOR platform in terms of how it could help SMEs. They require MULTISENSOR to provide them with the initial market research, the same information that PIMEC provides them. The key criteria were the following:

- Geographical proximity of the market
- GDP per capita
- Size of the sector
- Volume of imports
- Increase on the GDP
- Technological level
- Knowledge about the market
- Entrance barriers (customs)

- Specific regulations
- Competitors
- Price
- Corruption
- Fairs
- Associations
- Sector Magazines
- Distributors

From PIMEC, we then translated those requirements into more broad ones that suit the scope of the MULTISENSOR project, and proposed the following:

- Reachability between countries
- Economic situation
- Political situation
- Sector information
- Consumption Habits
- Competitors
- Associations
- Fairs and Magazines
- Distributors
- Product Characteristics
- Marketing Information

From these propositions, the internationalisation experts remarked that **reachability between countries** is very important in order to know how accessible or not a country is; **sector size information** is very important for them (it was pointed out that it is difficult to find the exact number); **product characteristics** would be a useful information regarding the SME itself, since they are the ones that will need to know how they will have to modify the product; and finally **consumption habits** were also remarked as important information, that they would complement with Product Characteristics information.

A big emphasis was put on MULTISENSOR providing a **comparative table** with the information from different markets, making it easy to compare and providing support with the decision making process. There was also a general agreement that there is the need of a person to interpret the results that MULTISENSOR will provide, as it will save a lot of time and make their decisions easier.

We also discussed the specific **legislations** of each country and market. They all agreed that it is an important information to know and consider when choosing a market, but that the information is not always accessible, and most of the times they get this by asking potential clients or distributors. We informed them that legislation is a difficult topic to summarise but there was a general agreement that

⇨ *If MULTISENSOR can only give them the parts of the legislation where their product is named, or at least the link of the right legislation, that will be already of great help.*

The internationalisation experts and the SMEs also had similar thoughts regarding **automatic language detection and translation** and **summarisation.** The internationalisation experts commented:

⇨ *"Summary of several articles or news could be time saving, but we also need the whole text to get the entire idea"*

⇨ *"Spanish is a must have language, since almost all the companies they work with don't have knowledge of other languages".*

Summing up, the feedback is that MULTISENSOR is a very interesting project for the internationalisation of SMEs and that it would be important if it could help them to access information regarding the **barriers to enter a country**, the **product characteristics** and the **sector information** together with **consumption habits**. Also, **summarisation** and **translation** of the texts to Spanish are a must for Spanish SMEs.

4.4 Core user requirements

Level 1 requirement	Description	Level 2 requirements	Priority[6]
Automatic search	Automatic search for each indicator and sub-indicator	1. The user can browse the indicators and sub-indicators 2. The system will automatically calculate initial values/reports for each indicator	1
User defined search	Search of predefined online sources according to user interests.	1. Continuous search profiles and one-time searches. 2. Search according to specific keywords, concepts (including hyponyms) and name entities (e.g. products). 3. Search in specific languages. 4. Search by topic. 5. The system must provide a list of the results, each of them represented by a small summary. 6. Item similarity search: the user should be able to select an article/item and retrieve relevant items.	1

[6] Meaning 1 = highest priority and 3 = lowest priority

Level 1 requirement	Description	Level 2 requirements	Priority[6]
Sentiment analysis	The software-based process of identifying the sentimentality (aka subjectivity) and the polarity of a piece of text, which reflects the author's opinion towards a specific subject. Sentiment analysis can be performed on a document-, paragraph-, or sentence-level, and can be specific to the entities, concepts, brands, or products mentioned in the text. The output is a set of scores in a predetermined scale.	1. Sentiment (polarity) analysis with regard to different entities (e.g. types of products, brands,) that are mentioned in one text. 2. Sentiment analysis based on keywords regarding a type of product	1
Image-based understanding	Understanding information extracted from an image	1. Extraction of high level concepts and events based on visual information and textual (if exists) 2. Enable visual similarity search between images focusing on product and logo search.	**1**
Automatic language detection and translation	Software-based language detection and translation of text documents from language to another.	3. Identification of language in texts/ audio files/videos 4. Translation of the textual content to language of choice	**1**
Correlation	The software-based process of combining (mainly textual) data with other related data, especially multimedia items and identification of correlations based on the detection of complementary, contradictory and similar content.	1. Identification of related social media content. 2. Suggestion of similar pieces of information from different sources (especially videos, images and multimedia items). 3. Detection of facts that are contradictory/complementary or similar mainly in different items.	2
Summarisation	Software-based process of reducing a text document in order to give the company a summary of the text.	1. Summarisation of long documents 2. Summarisation of social media	**1**
Comparison and decision support	A software-based process that allows the company to compare specific information about several countries	1. Creation of a comparative table with information about the different barriers of both countries compares 2. Calculation of risk based on the indicators for each country and presentation of the suggestion to the user.	**1**

Level 1 requirement	Description	Level 2 requirements	Priority[6]
Visualisations	Visualisation of MULTISENSOR analysis results and initial search configurations in an understandable and user-friendly way.	1. Network graphs of contributors and/or different content items. 2. Clickable tag clouds. 3. Pie charts, bar charts and line charts for different queries. 4. Map-based visualisation for country selection	2
GUI development	The MULTISENSOR interface should be easy to use and to understand	See mock-ups in this deliverable.	**1**

Table 3: Overview of basic requirements for Internationalisation Use Case

4.5 Summary and priorities

MULTISENSOR can be a powerful tool for Export Managers or internationalisation experts because it will help them reduce the amount of time they invest on accessing the information, correlating it and translating it.

In order for the platform to be fully helpful, the project should focus on **reachability of countries, product characteristics and sector information and consumption habits**.

Experts put a big emphasis on the **comparison** of information in a table or something similar, since it makes things easier for them to compare indicators of different markets and to make a sound decision.

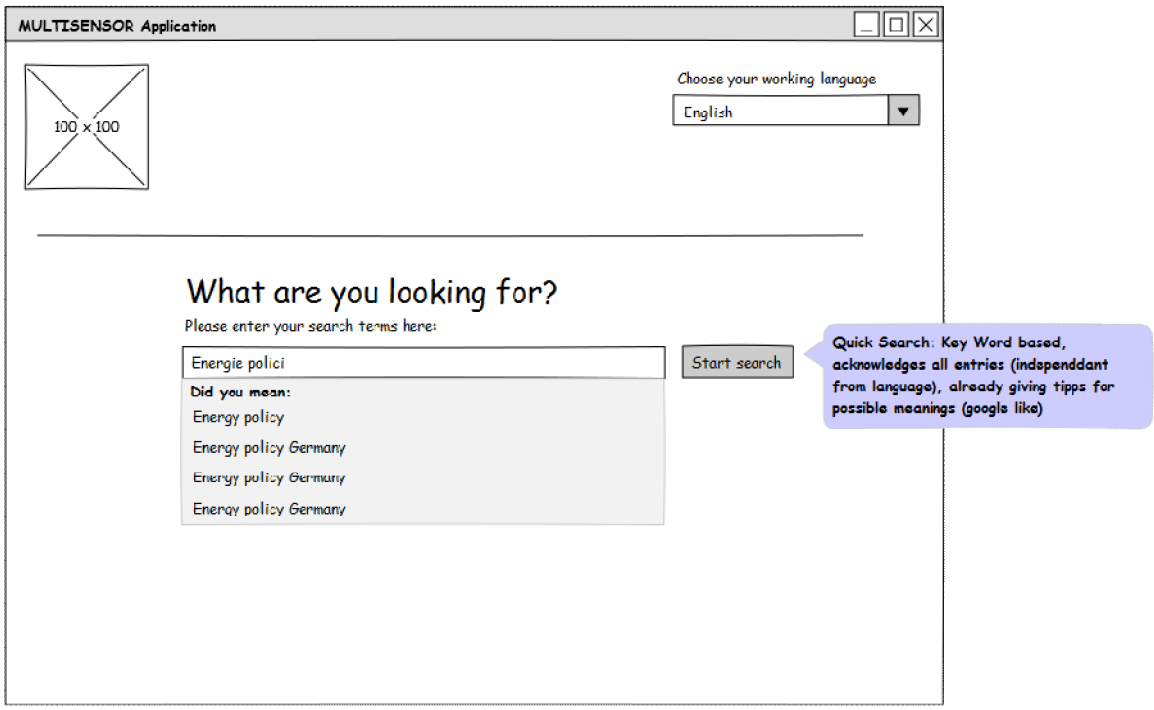
Because of saving time and especially because of the handicap that a lot of Spanish SMEs have in terms of languages (most SMEs only understand Spanish and Catalan, few of them also have knowledge of English and even less know also French), two very important requirements from our user case are the **summarisation** (to reduce the amount of information that the expert will need to read and study) and **automatic language detection and translation**.

5 USER INTERACTION: A POSSIBLE MULTISENSOR GUI

Based on the requirements and insights gathered through the market analysis and the expert interviews, a first draft of a possible user interface was sketched for each use case scenario. After preliminary discussions with the technical partners, these sketches were then transformed into the following wireframes, representing a first possible interface view but mainly reflecting the workflow of the use case. To avoid any doubts, these wireframes are much less an example of how the final system should look like but more a tool to demonstrate the typical workflow of a journalist, a media monitor or a business manager with regard to the issues that fall in the scope of the MULTISENSOR project.

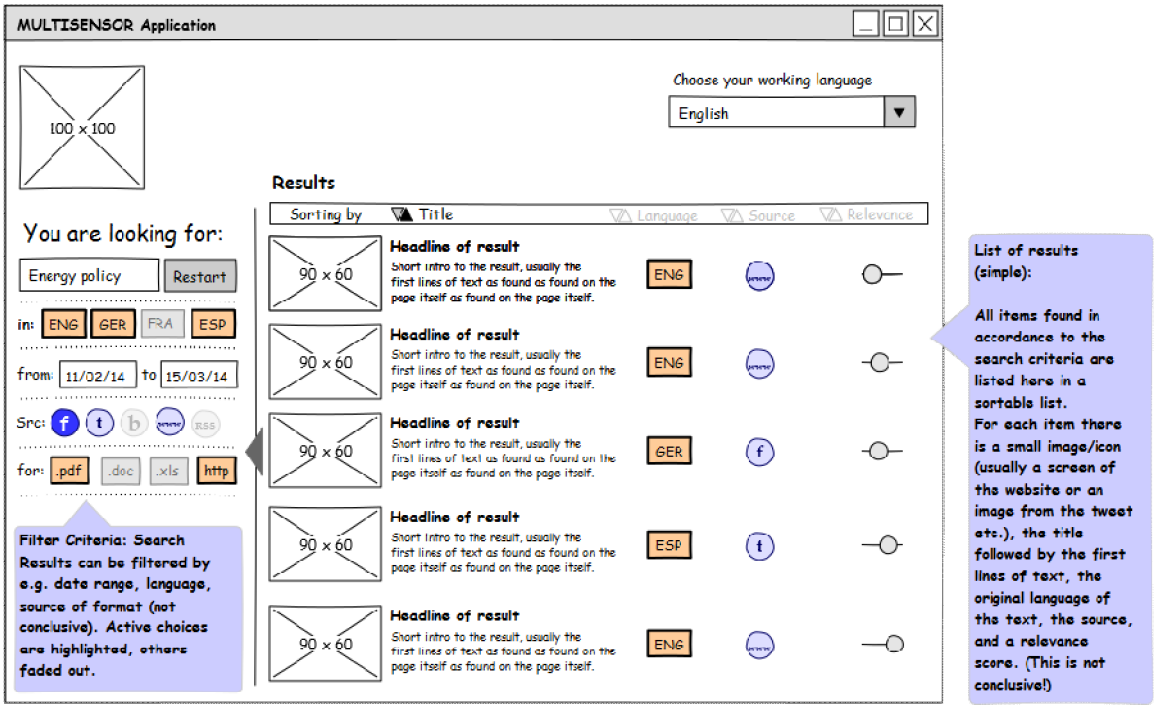
5.1 Journalism

As MULTISENSOR is supposed to assist journalists in their work, the most important aspect is the search function itself. The discussions made it clear, that the application should offer a very simple but smart start into the research process. Which is why for the journalistic use case, the application is based around a simple search, very similar to Google (see Wireframe 1).

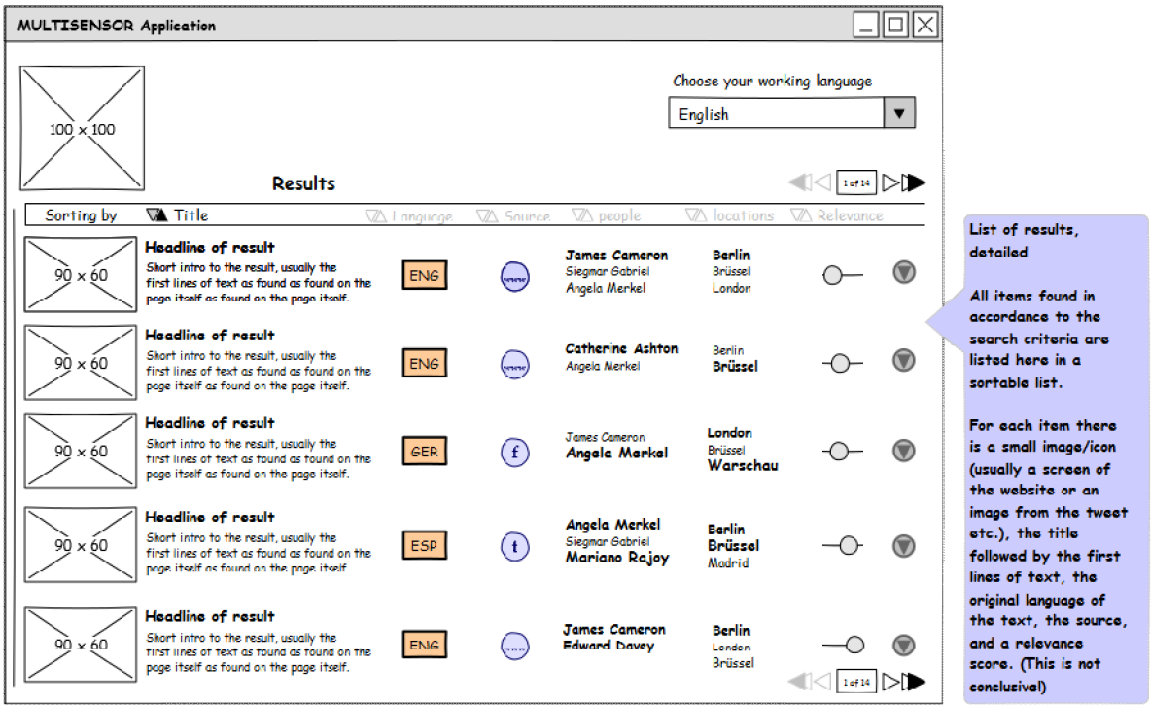


Wireframe 1: Search Screen for journalistic use case

While allowing to look for a single key word or a combination of such, this search also makes suggestions, correcting and/or complementing search terms. Once started, the search leads on to a list of results found within the available data. This list of results can then be sorted (e.g. by title or language) but also filtered further to refine the results (see Wireframe 2).



Wireframe 2: Search results and filter options

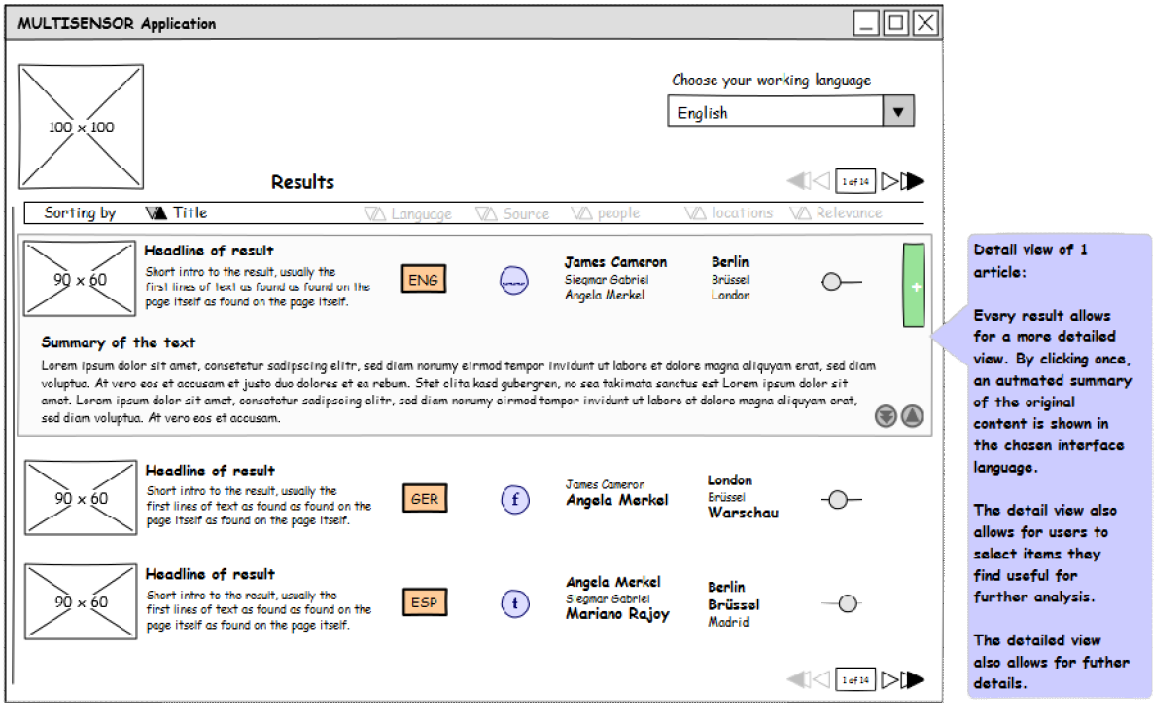


Wireframe 3: Search results with first analysis data

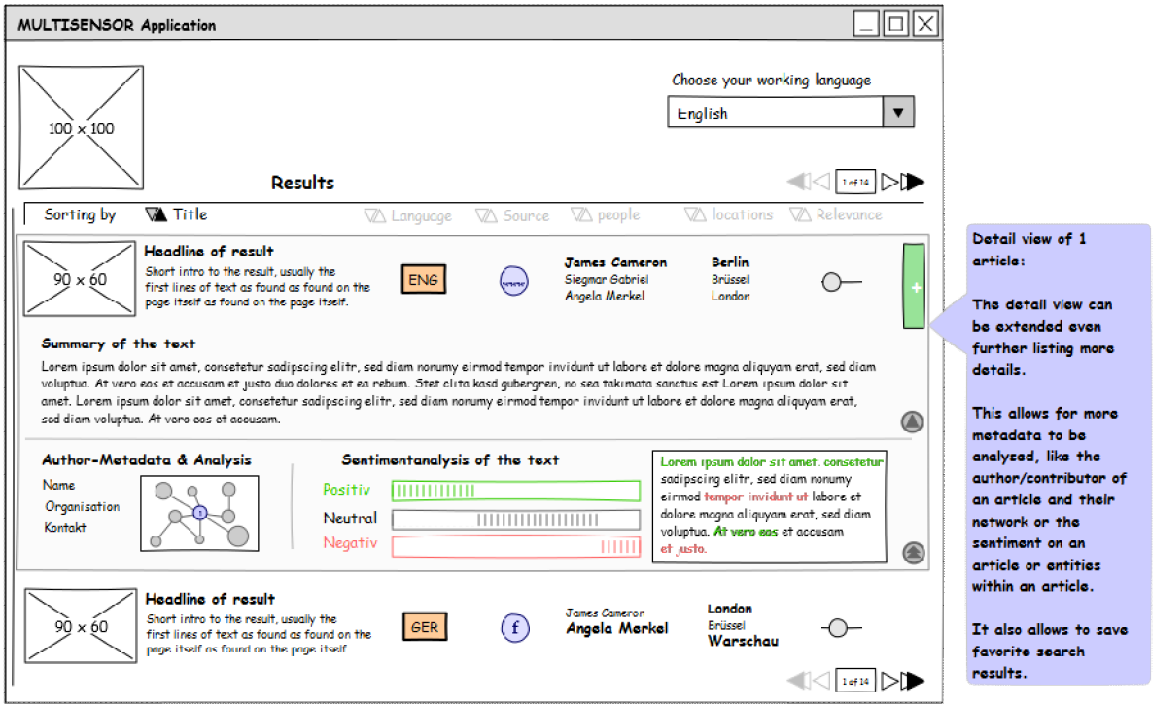
As for the previous as well as the following views of the application, the interface including the results is always shown in the language chosen by the user. The original language of the results is indicated (and also available as a filter and sorting option) in the list.

Once a user is satisfied with the list of results he can go on have a closer look at the data found according to his search criteria. By hiding away the filtering the interface gives more

room to do a preliminary analysis of the result list. In this view of the data, the user is presented with indications of the original language of the findings, as well as the sources, a first look at existing entities within each result (e.g. people or cities) or other possible small pieces of information connected to each single result (see Wireframe 3).



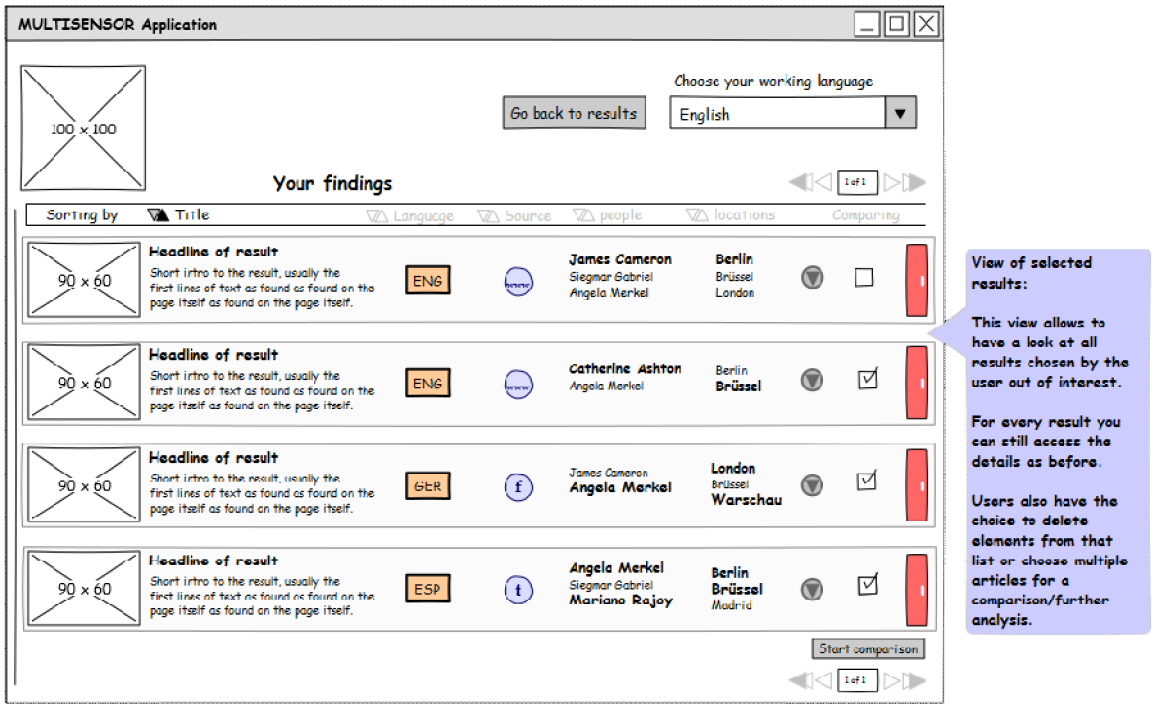
Wireframe 4: Opening the detail view on a search result



Wireframe 5: Second level of analysis details on a search result

In order to get more detailed data on one specific result, the user can then open up the data blog. This view offers the user an automated summary, again in the language of his choice (see Wireframe 4) and insight into metadata (e.g. Author, Network analysis) and analysis methods like sentiment scales or related articles or posts (see Wireframe 5).

Following the workflow of the journalistic use case the user can then choose results he is interested in for further usage by adding them to a personal findings page (see Wireframe 6).



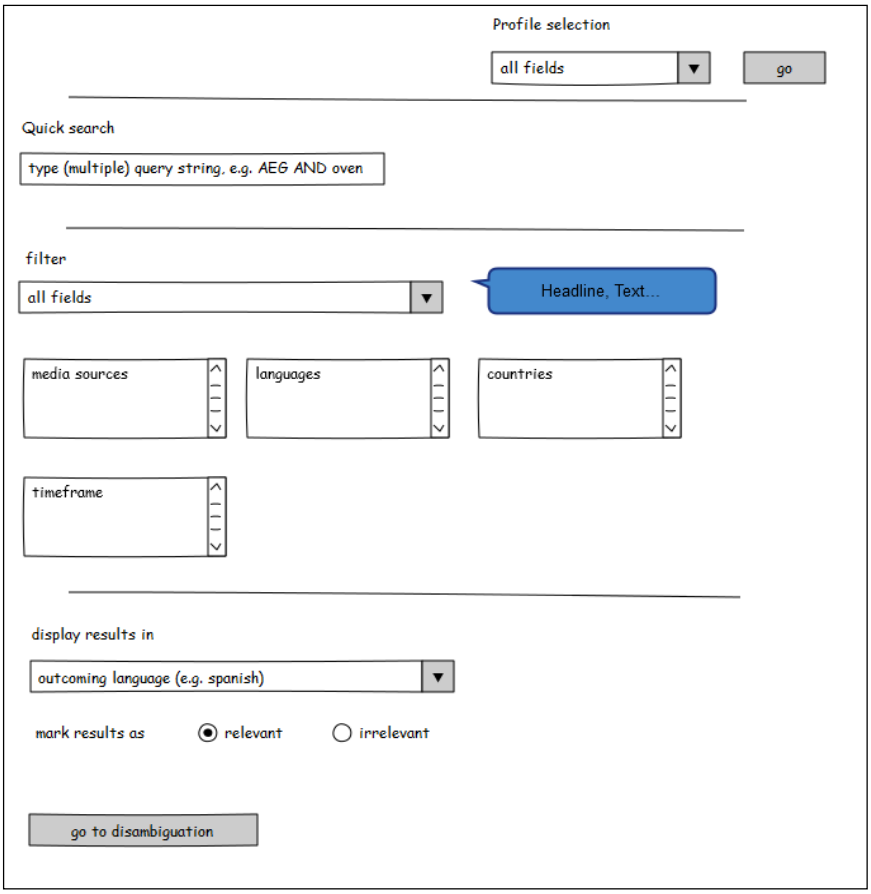
Wireframe 6: List of personal results, chosen by the user

Wireframe 7: Comparative Analysis of selected articles

To get even more insights the system should allow for a comparative analysis of the preferred results. This can be done by selecting articles from the personal view and starting the comparison process. This analysis will give similar feedback as the one before (sentiment, entities etc.) but on a group level, comparing a selection of results with each other (see Wireframe 7).

With this page, users can get a quick overview over tendencies within the results, helping them to better understand the content and choosing the materials they need for their work.

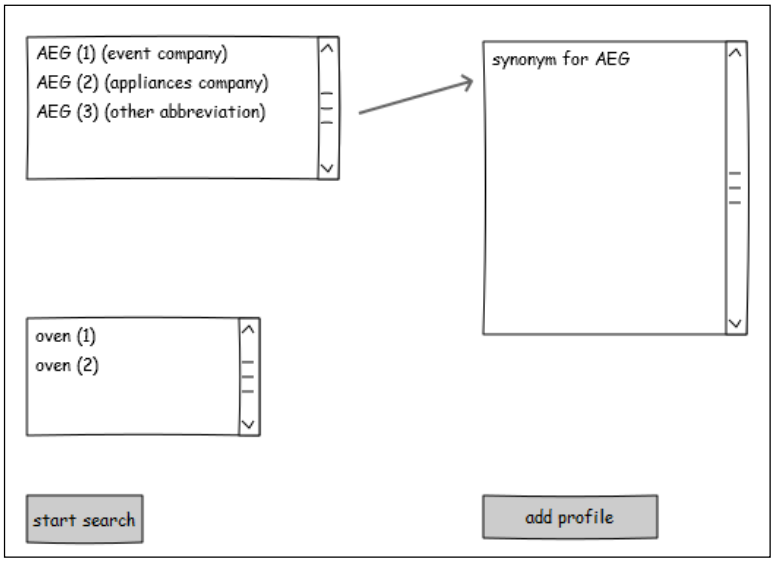
5.2 Media Monitoring



Wireframe 8: Search page for Media Monitoring Use Case

For the commercial media monitoring use case, the MULTISENSOR GUI should contain a search interface for string based or semantic searches that may either be one-time queries or stored profiles for running monitoring projects. The user should be able to define the time frame and precise media set in which he would like to search, taking into account several metadata as criteria. He will also define his output language of choice at an early stage. As the GUI should allow for the selection and validation of relevant content, the user may also choose to have his search results preconfigured as either relevant or irrelevant.

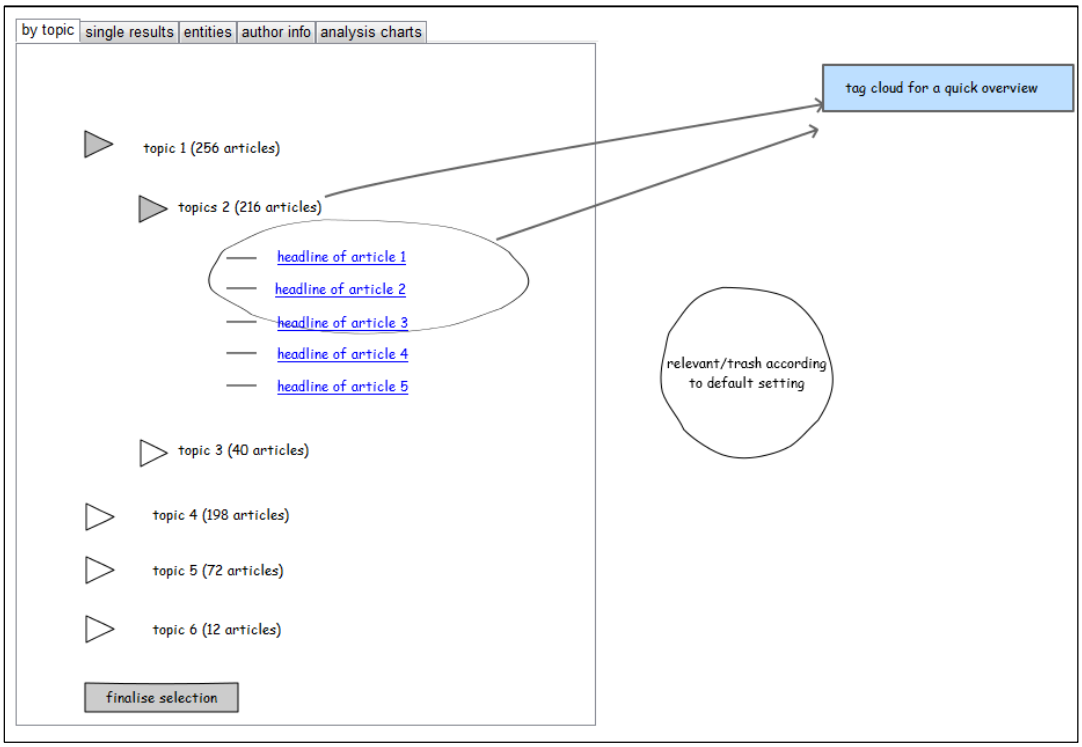
Before displaying the actual search results, MULTISENSOR should give the user the possibility to disambiguate his search terms and possibly offer him synonyms to include in the query. This will greatly help in reducing the amount of irrelevant search results.



Wireframe 9: Disambiguating search terms

After disambiguating, the user can either proceed to the results or store the search as a profile.

For the validation process, the search results should be displayed grouped by topics and, within these topics, clustered by similarity (e.g. identical articles should only have to be checked once). The user may drill down through the topics to the original articles. He also has the possibility to get a quick overview of the cluster's content by requesting a tag cloud. Individual articles and article clusters can be validated or discarded e.g. by drag and drop.

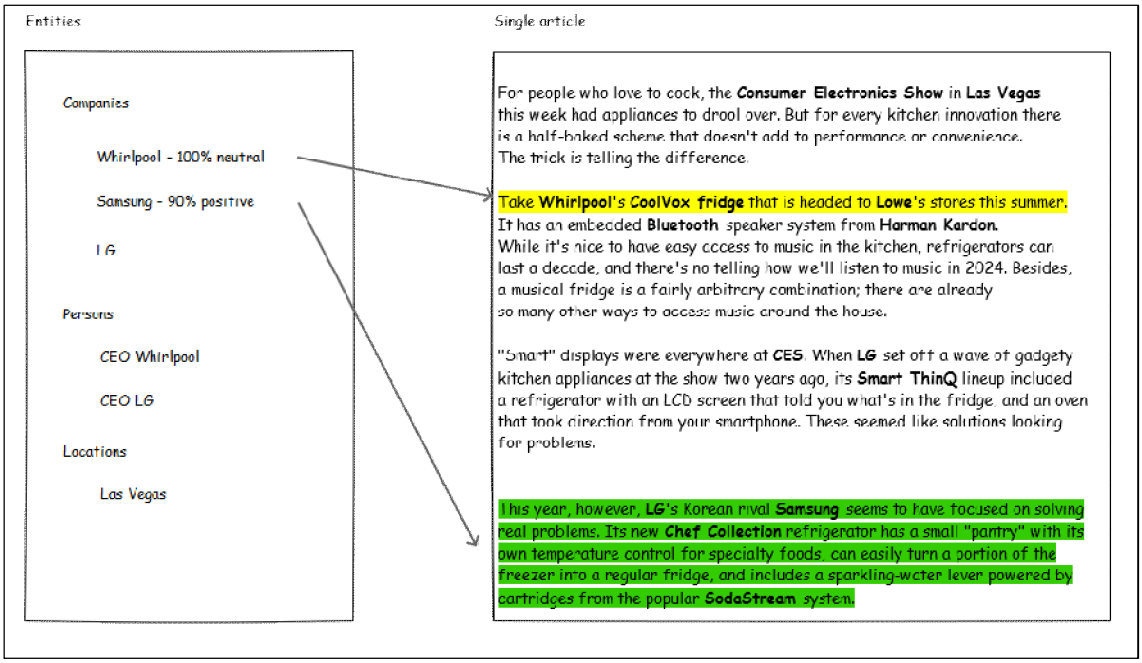


Wireframe 10: Displaying search results according to categories

Alternatively to the clustering by topics, the user may wish to go through and validate the individual articles. In this single article view, he will receive a list of articles with a short snippet and lists of included entities. Via buttons, the user will be able to compare translation and original article, get a short summary, information about the author or have the contained entities evaluated.

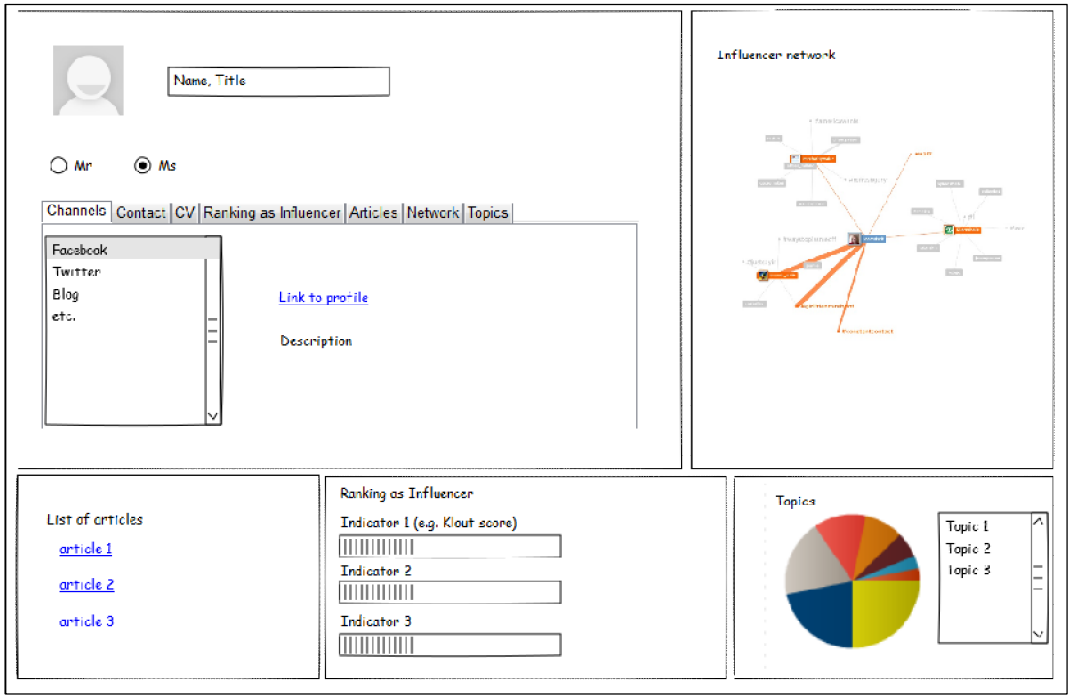
Wireframe 11: Displaying detailed views of results

When clicking on the *Evaluate Entities* button, the user will get a list of contained entities and the article text, in which the position of these entities can be seen. To get the tonality for an entity, the user can click on it in the list and will e.g. get the relevant sentences marked as positive, negative or neutral from the entity's point of view. The summed-up tonality may be displayed for the entity as a percentage score.



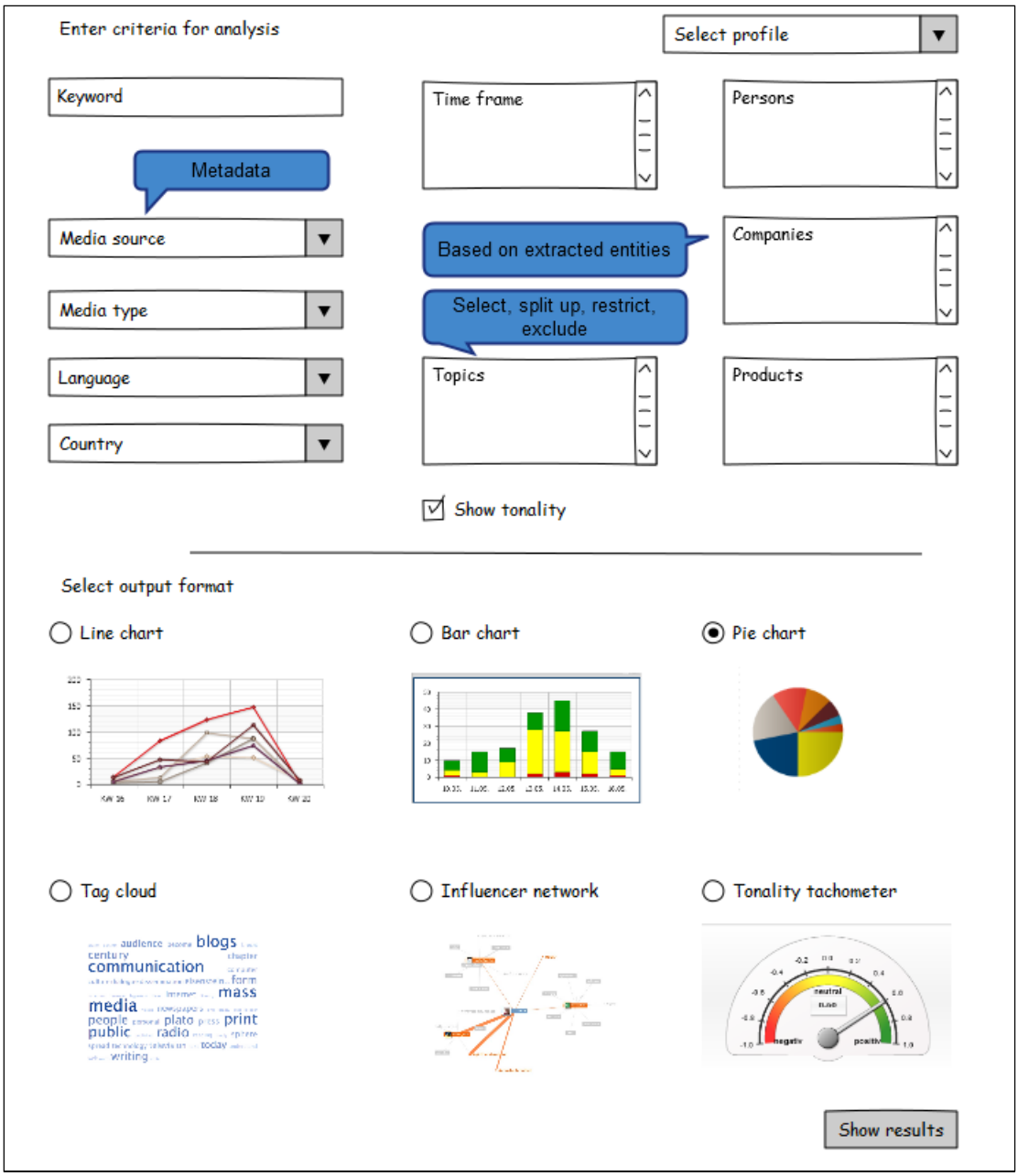
Wireframe 12: Detailed sentiment analysis per finding

Clicking on the *Show Author Info* button presents the user with a dossier about the author. This should depict a collection of previous articles as well as statistics about his activity across different channels and topics he writes about. Additional information on the author such as his CV etc. may be contained. On top, MULTISENSOR should be able to visualise the author's network and his importance as an influencer through the use of appropriate indicators that are to be defined.



Wireframe 13: Displaying Author information

Having selected the relevant news either through the single article or topic cluster view, the user can go to the analysis area and define the exact criteria for his desired analysis charts. He may select the search or profile he wants to use as data basis. The user should be able to choose the criteria he wants to have visualised from the metadata and lists of extracted named entities. For example, he selects all extracted products and splits them up by topic. The resulting output could e.g. be a stacked bar chart. Ideally, all criteria will be available for selection, splitting up, restriction and exclusion. There should also be a selection of possible output formats available (the picture below serves as example):

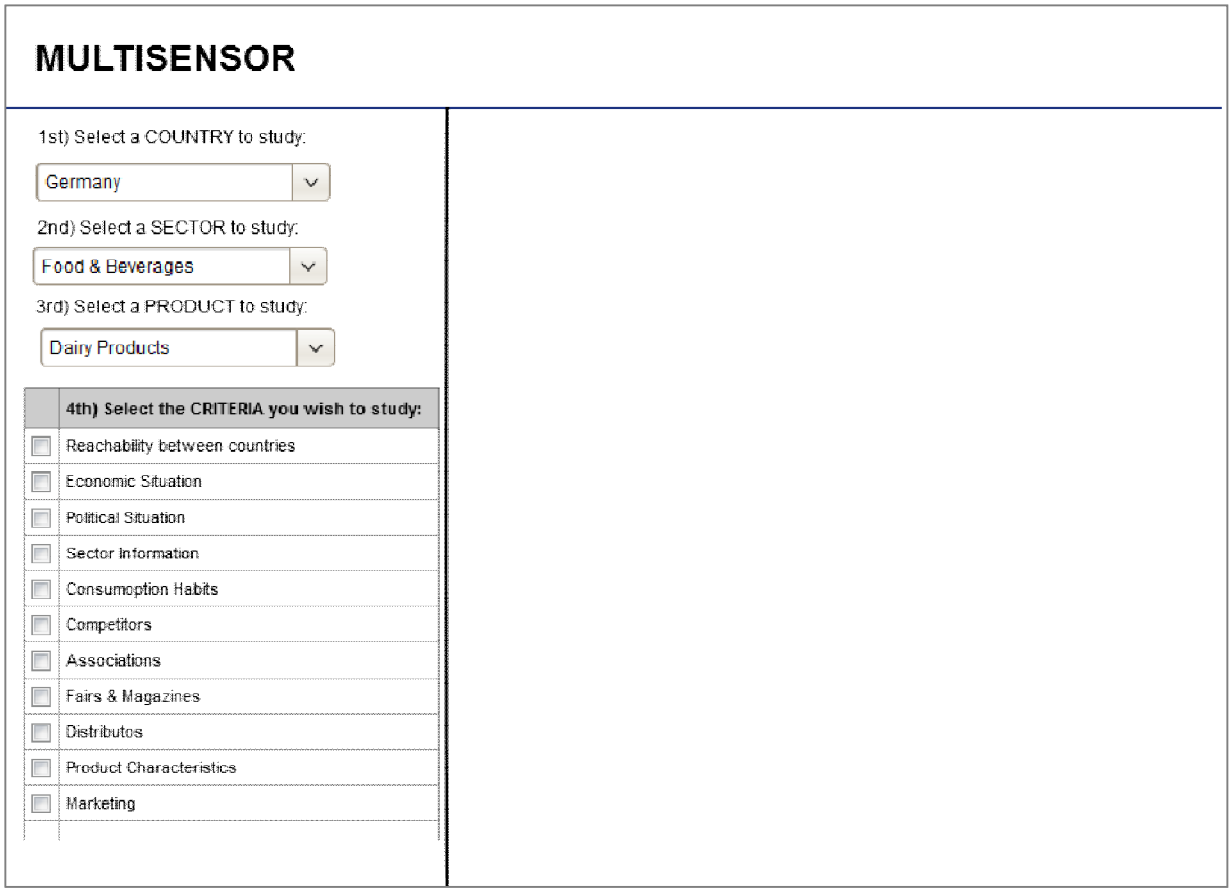


Wireframe 14: Detailed analysis of results

The generated charts need to be interactive and be linked to the article basis in the background. The summary functions should be accessible through the charts, e.g. it will be possible to create multiple article summaries for chart segments.

5.3 SME internationalisation

The user should be able to select from lists the country, the sector, and the product they wish to get information about. This information should stay selected through all the different pages of the platform and should be considered for the information in all the different pages of the platform. The user will also have a list of criteria that they can study. They will select the one they wish to get more information about and then the information regarding that criteria should appear on the right.



Wireframe 15: Start screen for SME internationalisation Use Case

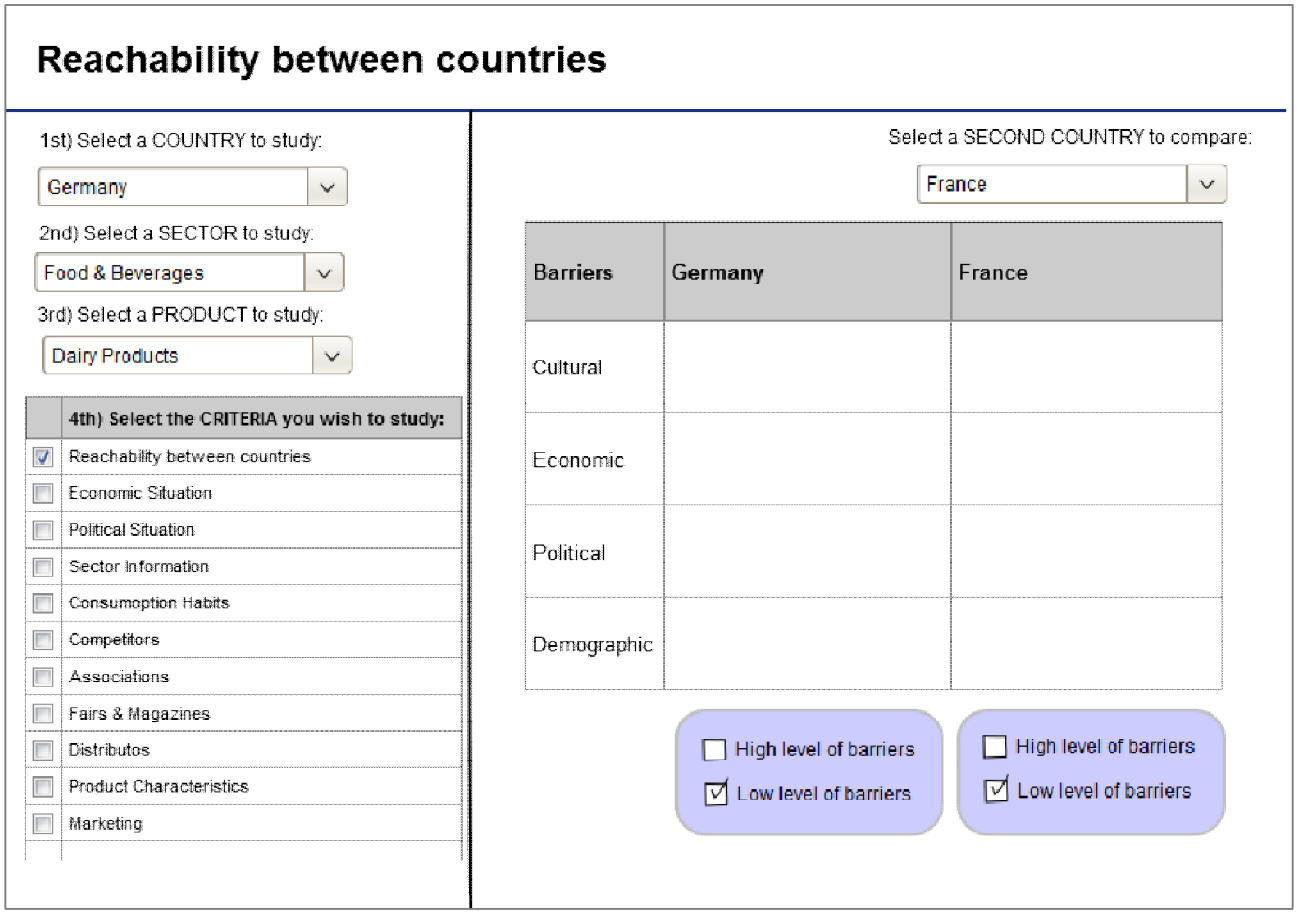
Reachability between countries

If they would like to study the reachability between countries, they will get a table. Then they will have the possibility to choose a second country to compare the information with. In the table they will be able to find information regarding the barriers that the selected countries have. Barriers should be divided into cultural, economic, political and demographic barriers, and the information should be presented in bullet points or small sentences.

- *Cultural Barriers:*
  This type refers to information that can be found on the countries WebPages about 'How to do business in Germany' (for example). This information will go from which language should be used, should we send gifts, should we dress very formally etc. An example of websites where this information can be found is
  http://businessculture.org/western-europe/business-culture-in-germany/business-

etiquette-in-germany/.
The information should be summarised and extracted from the text in order to get a short sentence or a small summary with the most important information to have into account.

- *Economic Barriers:*
  These barriers are more focused on GDP per capita, GDP growth, level of imports, how the economy is going to develop etc. These can be found on WebPages such as http://www.tradingeconomics.com/germany/gdp-growth or http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG.
- *Political Barriers:*
  This information is relevant in terms of the stability of the political situation of the market, recent news that might affect the export of any product or specifically of the SME's product, which political party is in charge, are there going to be any elections soon etc. All this information should be extracted from articles or news that are relevant in the market.
- *Demographic Barriers:*
  These barriers are the ones related to where the majority of the population is concentrated, in which areas we can find more potential customers, what is the average age of the country's population, life expectancy of the country's population etc. (if possible related to the products or sectors of choice).



Wireframe 16: Displaying reachability

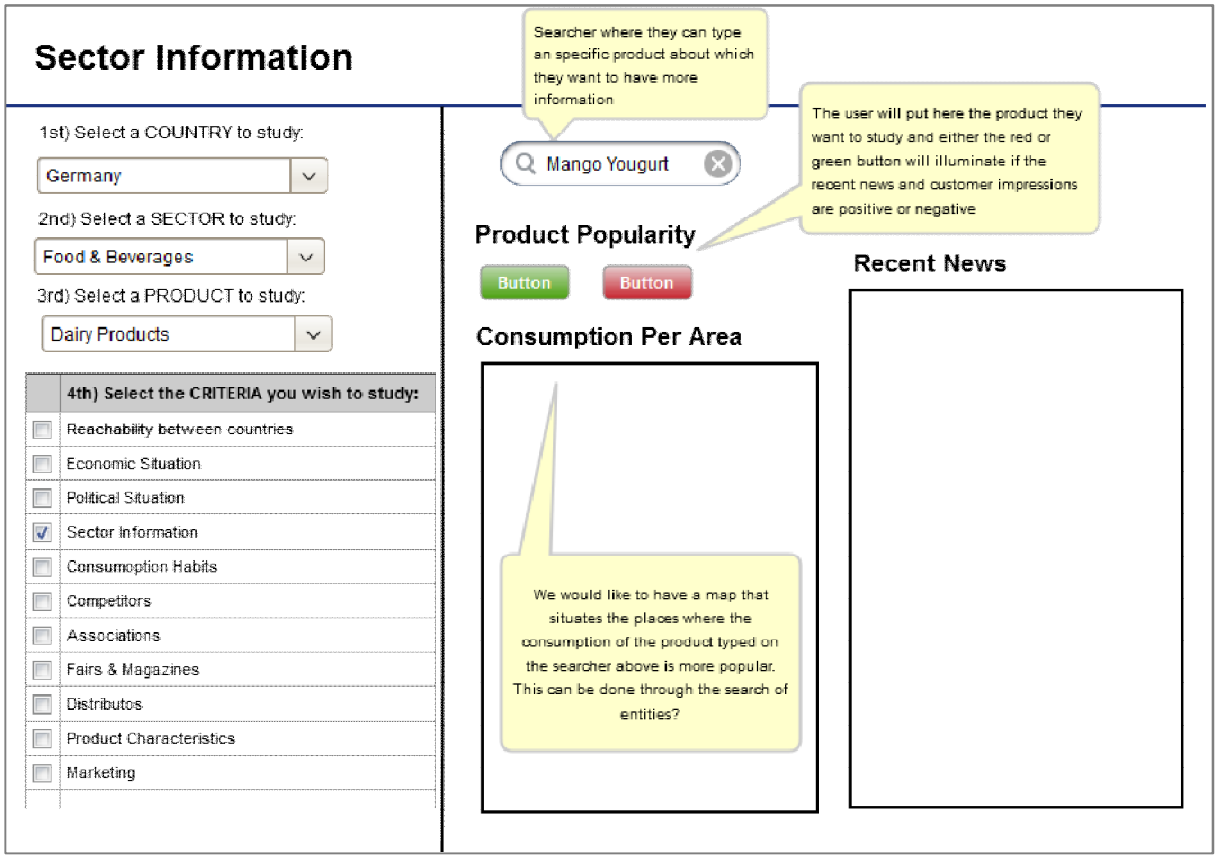
Once the information for the four barriers has been studied, the system should tick a box of 'high level of barriers' if the number of barriers is high, and 'low level of barriers' if the number of barriers is low. This should be done for both countries being studied separately.

If the user clicks on the criteria 'Sector Information', they will get three types of different information.

First, they will be able to select which specific product from the 'product family' they would like to have more information about (in this use case the family is 'Dairy Products' and the exact product is 'Mango Yoghurt'). Once the product is typed in, one of the buttons will light up depending on whether the recent comments on articles and social media regarding that specific product have been more positive or more negative. This information could also be presented as 'x' number of positive comments and 'x' number of negative comments, maybe a percentage or maybe as a velocity measurer. This is called **product popularity**.

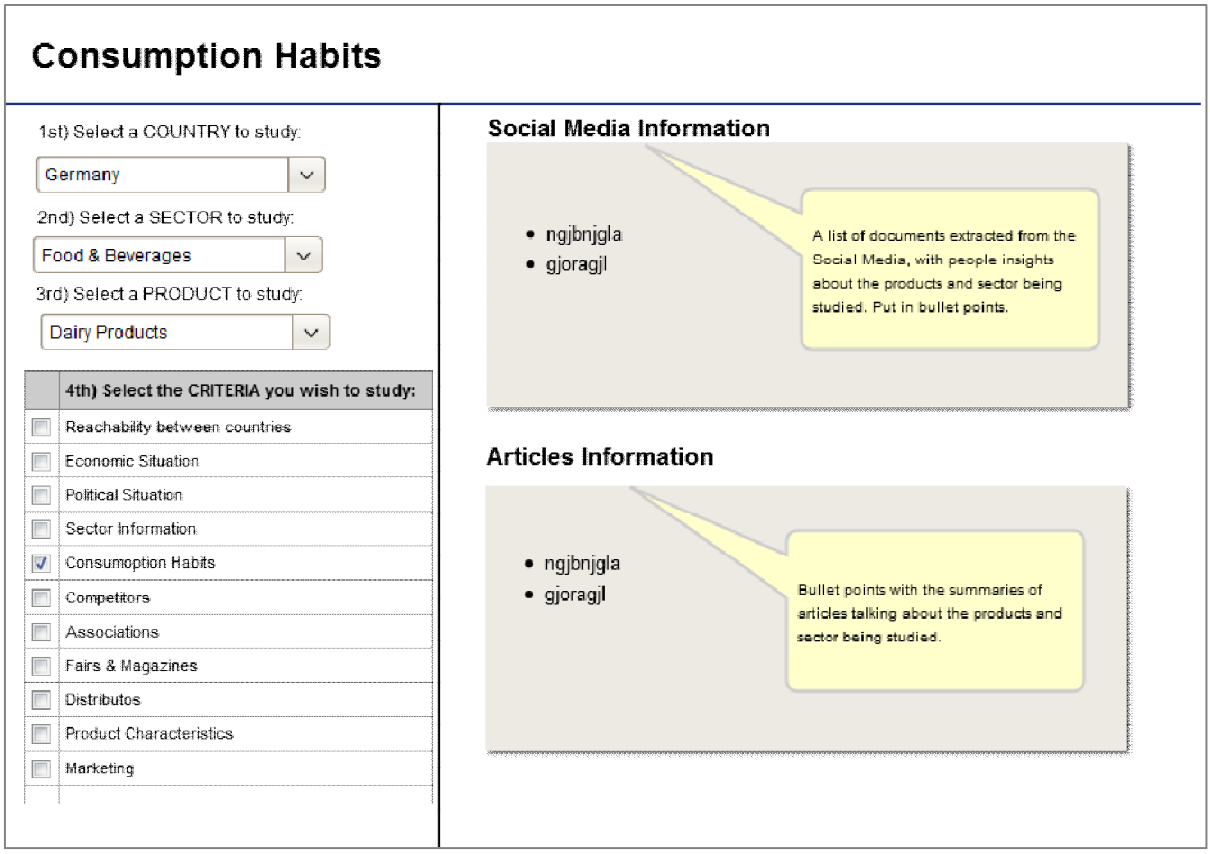
In the section **consumption per area**, the users will find a map with the areas of the selected country (in this example Germany) with more information on consumption of the chosen product. This information can be extracted from articles, compared and then put into an interactive map. If several articles name the area of Berlin as the area where there is higher consumption of Mango Yoghurts, the system will mark that area on the map.

Finally, the third part is a **recent news** area that would be a list of articles or news related to the chosen product. The articles should be shown as small summaries, with the option of choosing to read the complete document with a button 'Read more' or similar.



Wireframe 17: Displaying the sector information

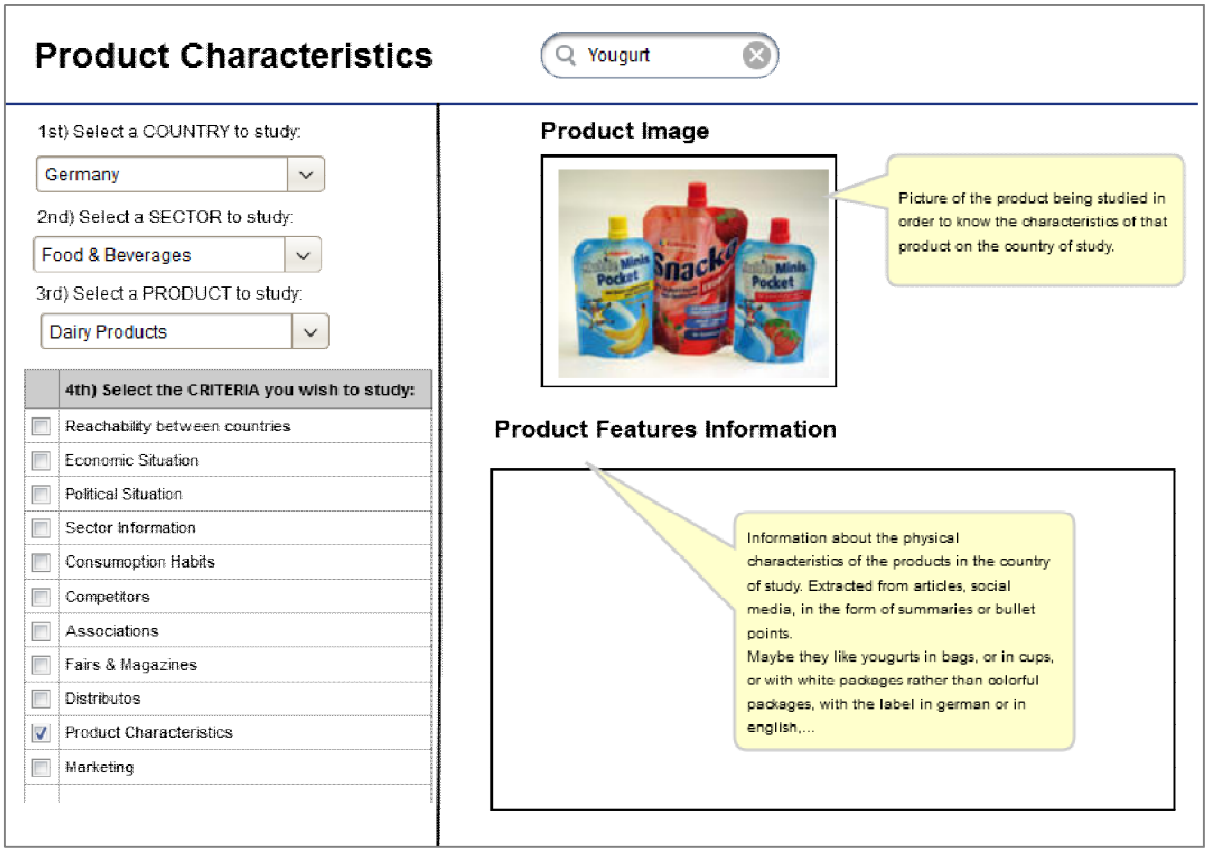
If they are interested on receiving information regarding consumption habits, users will find two different parts: **social media information** and **articles information**. In both parts, the user will find parts or small summaries of information regarding consumption habits of the general products (in this example 'Diary Products'). It would be interesting to find the information separated into those that come from Social Media and those that come from articles or news. The information should be presented as bullet points with the most important information.



Wireframe 18: Displaying consumption habits

Finally, if they wish to learn more about a specific product characteristic, users will find the following page. They will have to type in the general product they would like to study (i.e.: yoghurts, cheese, milk) and they will get different results. First they will find the **product image** with an image of the characteristics of the chosen product. This can be interesting because some countries prefer dairy products with more white colours that remind of milk, and other countries prefer more bright colours. It will also show whether yoghurt labels in Germany are usually in German or in English.

Secondly, users will find general **product features information**. This can be information extracted from news, articles or social media that refer to what is shown in the picture above. The information should be presented in bullet points or in very short summaries that makes it easy for them to see the information.



Wireframe 19: Product Characteristics

6 VALIDATION PLAN

The underlying evaluation methodology for user-centric trials has been extensively described in D8.1. We refer to this document and mainly to section 3.3 (from page 65). In this document, we will describe more specifically what we will practically do, e.g. **who** will validate the MULTISENSOR prototype **when** and **how**. In general, the evaluation of the first prototype will be mainly carried out by internal test persons as the system and its performance will be too immature to be tested by external experts. The following plan will be limited to the evaluation of the first prototype. Based on the experiences of this first evaluation we will develop the plan for further evaluation rounds which will be elaborated in the first evaluation report (D8.3)

6.1 Journalism

The development of MULTISENSOR user requirements for the journalistic use case is mainly based on two approaches: A competitive market analysis and in-depth expert-interviews with news professionals – either individually or in focus groups. Thus, both methods should be used for evaluating the MULTISENSOR system.

Existing tools that have been tested in the course of the market analysis will serve as a benchmark for the MULTISENSOR system and its prototypes. Through the competitive market analysis advantages and disadvantages of already existing tools have been identified, showing their shortcomings and thereby helping to define a market gap and a possible USP for the envisioned MULTISENSOR system in a professional news environment. Based on these insights, specific user requirements have been derived. The performance of the MULTISENSOR system should therefore be evaluated with regard to the performance of existing tools. The goal must be to compare MULTISENSOR's functionalities with regard to a possible additional value to the market.

Following a user centred design approach, the second phase of evaluation will focus on actual user-tests, involving real users from the target audience. The evaluation will consist of a mix of presentation, personal trials and interviews. Ideally in a one-on-one situation, the users will be presented with the MULTISENSOR project story. Starting with an overview of the project goals and its progress, users will be introduced to the necessary details and basic knowledge to perform the evaluation in an unbiased manner. Functionalities will be presented as well as their state of progress in as much detail as necessary to understand their purpose. This is to ensure a correct evaluation of the prototypes and project results, independent from the state of the MULTISENSOR application.

The actual user test will then be performed on the basis of a number of preset tasks. These tasks will be written in accordance with the project's state of progress, again taking into account that not all functionality may already be implemented or as perfect as aimed for at this stage. The order in which these tasks are exercised will follow a preset schedule. This includes standardisation in terms of number of questions, time for each step of the evaluation, support during the test etc. This method ensures comparable results over a number of test subjects and test sequences.

The progress and outcomes of these tests will be documented on the fly. All tests will be followed by individual interviews of the test persons based on a standardised questionnaire. Test users will be asked to comment on the tasks that they have performed, judging the system's functionalities and performance and also to express their overall assessment of the system. The latter questions are aiming at getting feedback on the tools usefulness for day-to-day professional work, its usability, and its economic potential for a bigger market.

The evaluation of the first prototype with regard to the journalistic use case will pursue the following steps:

What?	When?	Who?	How?
First prototype available	M18	-	-
First evaluation round	M19-20	Focus group with journalists and digital experts	The first MULTISENSOR prototype will be presented to a group of DW journalists and digital experts. They will be asked to validate the overall impression of the prototype, its main goals, interface and usability as well as specific functionalities that are already available. The focus group will also be asked to assess the performance of the first MULTISENSOR prototype in comparison with other already existing tools.
Second evaluation round	M19-20	One-to-one interviews with journalists and digital experts	A selected number of DW journalists and digital experts (as well as available SUG members) will be confronted with specific tasks that represent a near-real-world environment. After completion of these tasks they will be asked to validate the prototype's performance based on a standardised questionnaire.
First prototype evaluation report (D8.3)	M20	-	The first evaluation report will describe and analyse the evaluation process and results. Based on these findings it will eventually update user requirements and recommend certain improvements. It will also contain the plan for the second evaluation round (M25-26).

Table 4: Steps in first evaluation of the journalistic prototype

6.2 Commercial media monitoring

Evaluation for the second scenario of PUC1 will predominantly be qualitative and consist of user trials with several test scenarios. The targeted user groups in the media monitoring use case consist of two separate groups, namely a) employees working either as analysts or editors at media monitoring companies or b) managers and employees at client companies. The approach for the test scenarios should be different for both groups, depending on the evaluating user. The test routines for the client group should be focusing on the presentation of the data and the MULTISENSOR tool's overall usefulness. In this use case scenario, it is for the most part media monitoring professionals who are supposed to be involved with the handling of the data, so their test scenarios should cover workflow-related tasks in addition.

Additional to employees at pressrelations (who do not necessarily need to be involved in the use case creation and elicitation of requirements) who will perform trials with standardised predefined test routines, external testers, e.g. members of the user group with a media monitoring or client background, can be included in the trials once the system works stable. The content of the test routines needs to be mass-tailored to the components to be tested and their stage of development at the time of the test.

6.2.1 **Evaluation by media monitoring experts**

Analysts at pressrelations will carry out the following procedure in order to evaluate the MULTISENSOR tool: a complex, predefined query will be performed on the dataset available in the MULTISENSOR repository, covering all available data within a specific (short) timeframe. In order to evaluate the results returned by the MULTISENSOR tool, a gold standard set will be created manually in advance, providing the relevant results for the specific query, including additional information like topic clustering, language detection, named entities, sentiment analysis, and author/contributor information. The gold standard set will be created semi-automatically by using pressrelations' NewsRadar system. The gold standard set will function as the baseline for automatic as well as manual evaluation of the particular MULTISENSOR tools. Whenever it is feasible, testing will be carried out automatically and on a quantitative basis, but in order to determine whether MULTISENSOR meets the needs for the media monitoring workflow, certain components need to be evaluated manually.

Requirement	Automated	Manual
Topic detection and clustering		X
Language detection		X
Keyword disambiguation		X
Named entities extraction	x	
Automatic summarisation		X
Machine translation		X
Sentiment analysis	x	
Contributor analysis	x	
Overall usefulness		X

Table 5: Evaluation processes for the separate components

The query the analysts will perform contains simple keywords (strings) as well as semantic concepts that are part of the knowledge base. Additionally, the search will be specified with respect to language, country, media sources, and the time frame in which the items were generated. The search results in the test set will be returned as topic clusters which will be checked for their validity and coherence by the analysts. Depending on the granularity of the topic model, the MULTISENSOR tool should return comparable thematic clusters that have been established manually in the gold standard set. As a side effect of the evaluation of the topic model, the disambiguation of keywords can be validated in the same step. Since distinct readings of a keyword are prone to occur in distinct topics, the analysts can evaluate whether all identified topics are instantiations of the semantic concept in question, rather

than of keywords in a homonymic relationship. The same holds for the automatic language detection functionality. Since topics are usually based on word co-occurrence, items in the same language appear in the same topics, so that analysts can easily identify items that have been falsely analysed.

On article/item level, the analysts will compare the named entities that have been extracted automatically in the test set to those that have been coded manually in the gold standard set. Alternatively, the evaluation of the named entities per news item can also be carried out automatically, using a simple diff function to compare the automatic to the manual lists. The performance of the MULTISENSOR named entities extraction tool will be measured by calculating precision and recall against the gold standard set.

For two functionalities of the MULTISENSOR tool there will be no gold standard solutions: 1) the automatic summaries as well as 2) the translations will be evaluated individually, since quantitative validation of the results tends to be difficult, and qualitative judgments of machine translations as well as automatic summaries are inherently a question of taste. Since the overall goal of the summaries and translations within the MULTISENSOR media monitoring user part is to assist media monitors in their decision making process, the main criteria for the evaluation of both functions will be fidelity and intelligibility, both of which are hard to measure automatically. 'Fidelity' refers to the amount of (original) information that is preserved in the translation/summary; 'intelligibility' refers to how understandable the resulting text is. In order to ensure a high inter-evaluator agreement, a list of key criteria for both summaries and translations will be developed in advance. Since pressrelations has (native) speakers of all five languages covered in MULTISENSOR in house, the evaluation of summaries and translations can be done individually after a short training period.

The quality of the sentiment analysis within the MULTISENSOR tool will again be measured against a gold standard annotation. For all the items that are considered relevant in the test set, the tonality of specific sentences and paragraphs will be marked on a scale. Each sentiment is assigned from the point of view of a particular named entity. For the test set, this process will be carried out automatically, for the gold standard set this task is carried out manually. The results of the automatic sentiment analysis in the test set will be compared to the manual analyses in the gold standard set, and performance will be measured by calculating precision and recall.

In order to evaluate the contributor analysis functionality, all available information about the author of an item will be manually gathered for the gold standard set. The results of the automated MULTISENSOR analysis will be compared to the handcrafted author database.

On top of the quantitative and qualitative evaluation of the particular functionalities of the MULTISENSOR tool, the testers' satisfaction with MULTISENSOR and their assessment of the tool's usability and overall quality of the results will be recorded through questionnaires and interviews. For the near final product, important quantitative aspects such as time consumption for a given task should be tracked and compared to existing workflows.

What?	When?	Who?	How?
Preparation of reference data set and questionnaire	M16-17	PR staff	Manual annotation of gold standard set using NewsRadar. Creation of an appropriate catalogue of questions for the assessment of the system's overall usability.
Evaluation round	M18	PR staff	Media monitoring employees will test the overall usability of the MULTISENSOR tool and compare the data output for a specified query with the gold standard data.
Creation of the First prototype evaluation report (D8.3)	M19-20	PR staff	The results of the evaluation round will be assessed in the deliverable D8.3 and possibly lead to an amendment of the user requirements.

Table 6: Steps in evaluating the media monitoring prototype

6.2.2 **Evaluation by media monitoring clients**

The second part of the evaluation of MULTISENSOR will be carried out by managers or employees of media monitoring clients. Since they are not involved in the process of gathering and analysing data, the evaluation will focus on MULTISENSOR's general usefulness and the presentation of the results. External users will not be involved before the prototype runs smoothly.

6.3 SME internationalisation

The development of MULTISENSOR user requirements for SMEs use case will be mainly qualitative, based on two separate groups: a) companies interested in exporting their products and b) export managers and export related workers.

First of all we look up all the sources for making a market analysis to identify advantages and disadvantages of the tools that the companies and the export managers use to open a new market. Once we evaluate all this information we can use it for the focus group.

The first group is about companies, which have a deep knowledge about their product and sector, so they know where to go and they have tailored their potential markets. In order to identify companies` needs, inwards and outwards, focus groups will be carried out. The main goal in these focus groups is to find out the current situation of the company, which obstacles, barriers or hazard these companies could find either in the short and long run. The performance of this group is to realise which are the needs to focus on and also to create additional value which will be an effective tool in order to assist and assess companies to implement their international strategy.

The second group, an important part of the trial, brings its experience and knowledge about internalisation and export to companies. They assess the companies where they may go, where they should go, which are the barriers and obstacles of each country, which are the possible channels of entry and whom they should look for (e.g., wholesalers, distributors). The extensive interviews and the focus group with the export managers will have constituted the core source of user requirements; consequently, user trials with companies should also be the core of the evaluation process.

In this use case scenario, it is for the most part export manager professionals who are supposed to be involved with the handling of the information, so their test scenarios should cover workflow-related tasks in addition.

Also company testers can be included in the trials once the system works stably. The content of the test routines needs to be mass-tailored to the components to be tested and their stage of development at the time of the test.

The testers' satisfaction with MULTISENSOR and their assessment of the usability of the tools and overall quality of the results will be recorded through questionnaires and interviews. For the near final product, important quantitative aspects such as time consumption for a given task should be tracked and compared to existing workflows.

After each cycle of user trials, the testers' feedback will be analysed and the results will be set into correlation to existing tools and workflows as specified in the competitive analysis that will be performed for D8.2.

What?	When?	Who?	How?
First prototype available	M18	-	-
First evaluation round	M19-20	Focus group with SMEs and Internationalisation Experts	The first MULTISENSOR prototype will be presented to a group of SMEs' export managers and internationalisation experts. They will be asked to validate the overall impression of the prototype, its usability and if the presentation of the results and the system's functionalities are helpful for them. The focus group will be asked to try and assess the performance of the prototype in comparison with the procedures that they normally use on their day to day work.
Second evaluation round	M19-20	Interviews with Internationalisation Experts	A selected number of Internationalisation Experts will be gathered and asked to solve several problems they might face when helping an SME internationalize (e.g. search of consumption habits) using the MULTISENSOR prototype. After performing this task they will be asked to fill in a questionnaire explaining their experience, their opinions and things they find can be improved in terms of usability and presentation of results.
First prototype evaluation report (D8.3)	M20	-	The first evaluation report will describe and analyse the evaluation process and results. Based on these findings it will eventually update user requirements and recommend certain improvements. It will also contain the plan for the second evaluation round (M25-26).

Table 7: Steps in evaluating the internationalisation prototype

7 CONCLUSIONS

In this report, we have considered the use of information and data from heterogeneous and multilingual sources by journalists, commercial media monitors and business managers. We have identified and discussed the needs of these groups in depth, identifying shared and distinctive features. We have also identified "competitor" systems and considered their strengths and weaknesses in detail, in order to help define our own system requirements. We have developed a first validation plan and demonstrated possible user interactions by creating basic user interfaces for the different use case scenarios.

This report will be used to guide the development of the prototype MULTISENSOR system and will be updated in a series of evaluation reports (D8.3, D8.4 and D8.5). Independent of these evaluation reports, the process of defining user requirements has proved the need for an iterative R&D approach that includes permanent exchange between technical partners and user partners. The goal of all this work is to help ensure that the project remains focussed on user needs by developing, refining and sharing the user's requirements.

A APPENDICES

A.1 Market analysis and interviews

A.1.1 Pilot Use case 1: Journalism

Market analysis

For the market analysis the Deutsche Welle used different ways to track down possible tools with similar functions. It relied on the expertise of its journalistic network, interviewing its members about what tools they currently use to search, monitor and analyse online content. It also used several social media channels to tap into the knowledge base there and did general online research on possible tools according to the main focus areas of MULTISENSOR.

As a result a list of the findings was created dividing the individual tools into different categories, according to their main functionality, making it easier to get a quick overview of the different tools. Every category is represented by one or more exemplary tool, similar tools are listed below. For each of these exemplary tools there is its name and a short description, its web-address, an overview of its functionalities and its relevance for MULTISENSOR.

Due to the constant growing number of online tools this market analysis can only highlight the functions covered by these tools. It is not considered a concluded list.

Social News Aggregator

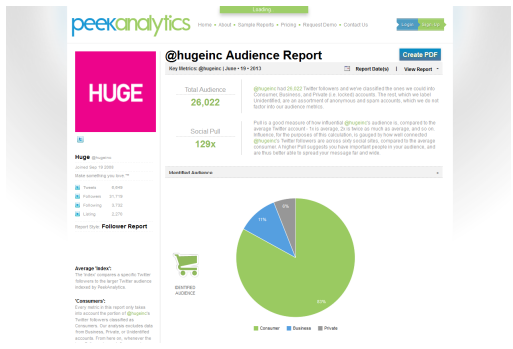
The Category Social News Aggregator contains tools like virato, tame.it or rivva that are meant to give users a quick overview of what is currently discussed on the web and what topics are trending. The tools all rely on the number of shares a topic (respectively an article) received in order to come up with a ranking. The majority of these tools are not focused on one particular social network.

Title	**Virato – social news**
URL	http://www.virato.de/
Description	Virato provides prioritized lists of online news. Articles are picked from official news sources, blogs and social networks. The ranking is done according to the number of likes, Shares and other social rating factors
Functionality	fully automatic ranking, choice of news categories
Relevance for MULTISENSOR	• the data aggregation itself • the social ranking algorithms • the visual interface of how the lists are presented
Similar tools/sites	• http://www.10000flies.de/ • http://sulia.com • http://www.socialmention.com/ • http://tame.it/ • http://rivva.de/ • http://curator.buzzrank.de/

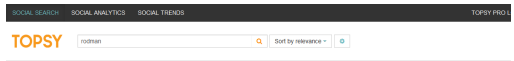
Twitter Network Search and Analyses

This category is focusing on Twitter as a source, giving the user control over their twitter channels. Peek Analytics and the other two mentioned tools help identifying influencers and popular messages by providing detailed analysis of ones own tweets. The second set surrounding topsy focuses more on search functionality within the twitter network, also offering filter-options and sentiment analysis of the content found.

Title	**Peek Analytics**
URL	http://www.peekanalytics.com/
Description	Peek Analytics is a tool used for the measurement of Twitter audiences. It helps bring out numeric details about followers, followed, social influence and messages.
Functionality	Analysis per account, keywords or person.
Relevance for MULTISENSOR	• the analysis of network structures • the social influence algorithm • the visual interface of how the numbers are presented
Similar tools/sites	• http://mentionmapp.com/ • http://klout.com/home



Title	**Topsy**
URL	http://topsy.com
Description	Topsy is a social media search engine, focusing mainly on Twitter, allowing for a

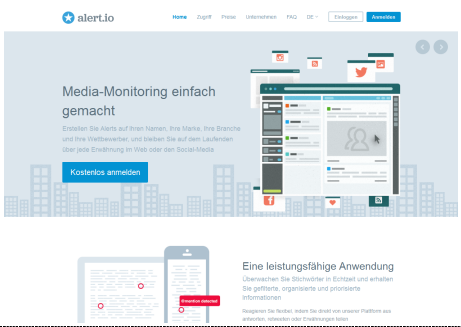


	keyword (or hashtag) based tweet search. In addition the tool offers a few analysis functions to filter and sort the messages further, like timeframes, languages, Sender.
Functionality	Search engine, filter options (time, language, source), sentiment analysis, trend analysis on most current tweets
Relevance for MULTISENSOR	• filter options for time, source and language • Visualisation of current trending topics • the visual interface of how the results are listed
Similar tools/sites	• Vox Event Analytics: http://sm.rutgers.edu/vox/event/#s=&f=0&i=0&v=0&so=1&g=0

Online Media Monitoring Services

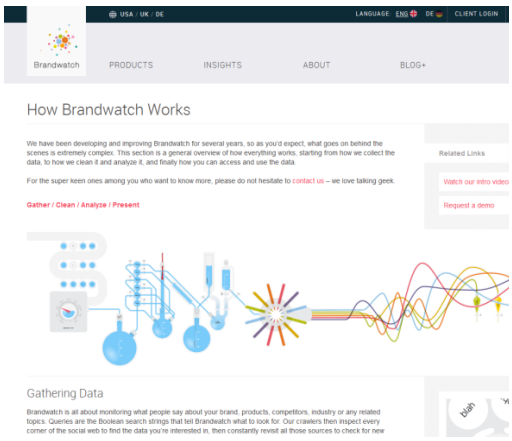
This category lists tools that have a stronger focus on classic media monitoring tasks, but have been adjusted to the online world (web-monitoring/social channels). The first set with alert.io is similar to the Google alert system, offering more options and a focus on social media. The second set has a stronger business focus. The third set (Newsexplorer) takes a closer look at traditional media sources (major newspapers like the New York Times) and offers language filters and entity extraction.

Title	**alert.io**
URL	https://de.alert.io/
Description	alert.io (formerly „mention") is a Google alerts like web-tool. The system monitors the web (blogs, social media, news outlets …) according to keywords entered. Results are displayed via a user dashboard. The user can scroll through the results, reply to messages or collaborate with colleagues.
Functionality	Web monitoring via keywords, sentiment-analysis, supporting several languages
Relevance for MULTISENSOR	• Aggregation of news from different online sources • Visualisation of results in a multifunctional dashboard • sentiment analysis • clear Visualisation of sources • the visual interface of how the results are listed
Similar tools/sites	• http://sproutsocial.com/ • http://www.brandwatch.com/key-features/
Title	**Newsexplorer**
URL	http://emm.newsexplorer.eu/NewsExplorer/home/en/latest.html
Description	Newsexplorer is an online media monitoring service. The focus is on traditional news outlets like BBC or Euronews. The system lists the most current news found in the given sources, but also lets you filter the news by language, region or date while offering a search as well as a few statistics about the results.



Functionality	Live media monitoring of online media, keyword search, event detection, basic analysis functionality, supporting several languages
Relevance for MULTISENSOR	• Multilanguage feature • Event detection • Entity extraction
Similar tools/sites	

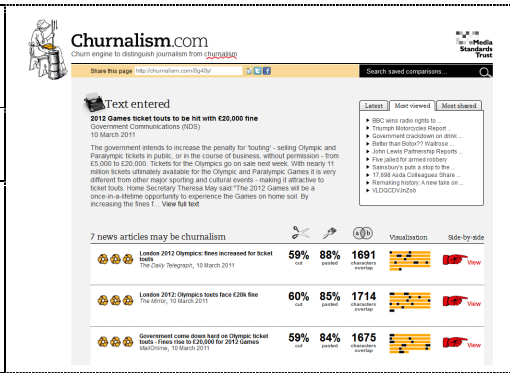
Title	Brandwatch
URL	http://www.brandwatch.com
Description	Brandwatch is an online media monitoring tool, similar to the aforementioned tool alert.io. The difference is that besides the keyword search, Brandwatch offers some algorithms to analyse and enrich the findings, allowing for a more thorough analysis
Functionality	Complete, location based services, sentiment & topic analysis
Relevance for MULTISENSOR	• Algorithms to analysis and enrich findings • Visual Interface
Similar tools/sites	• http://www.attensity.com/home/ - large number of input channels & analysis options • http://www.quintly.com/



Text analysis, extraction and comparison

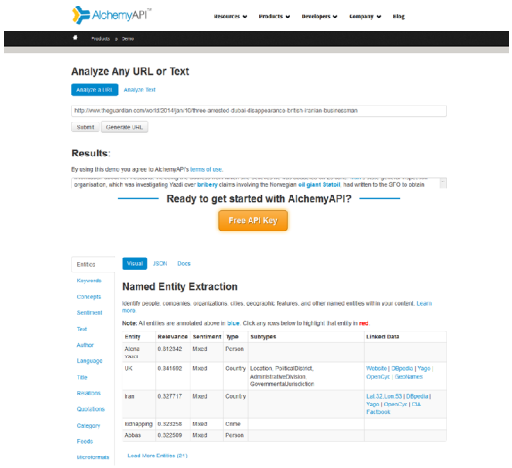
In this category the focus is on text analysis more than research functionality. All four sets present tools that allow Natural Language Processing (NLP), comparing of texts as a whole or as fragments and extraction of entities for further analysis.

Title	**Churnalism**
URL	http://churnalism.com/
Description	Tool that helps analysing articles to see whether they have elements or even paragraphs copied from other texts online.
Functionality	Comparison of text elements and analysis of degree similarity
Relevance for MULTISENSOR	• Text analysis and comparison • the Visualisation of the results
Similar tools/sites	• http://winmerge.org/

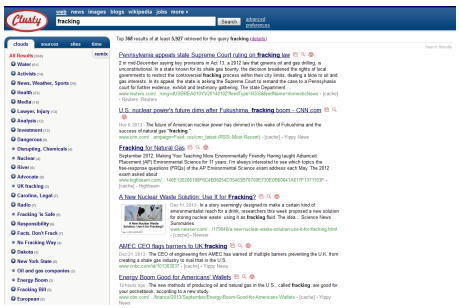


Title	**Open Calais**
URL	http://www.opencalais.com
Description	Open Calais is an NLP-framework (Natural Language Processing) to analyse text-based documents. The texts are fragmented and semantic entities and other metadata is extracted (e.g. names, places, dates, amounts, etc.)
Functionality	Entity Extraction
Relevance for MULTISENSOR	• metadata extraction and Visualisation
Similar tools/sites	• Semantria: https://semantria.com/demo

Title	**Alchemy API**
URL	http://www.alchemyapi.com
Description	Alchemy API is a software that offers text extraction and analysis via API
Functionality	Text analysis on entity, topic, people, dates, sentiment basis
Relevance for MULTISENSOR	• Entity Extraction (people, companies and organisations mentioned in text) • Keyword Extraction • Sentiment Analysis • Relation Extraction (detect important events and signals between entities) • Language Detection • Author Extraction
Similar tools/sites	• http://www.connexor.com • http://www.basistech.com/ • http://www.datumbox.com/ • http://gate.ac.uk/ (main contributors are U Sheffield and Ontotext) • http://www.ontotext.com/kim (uses GATE)

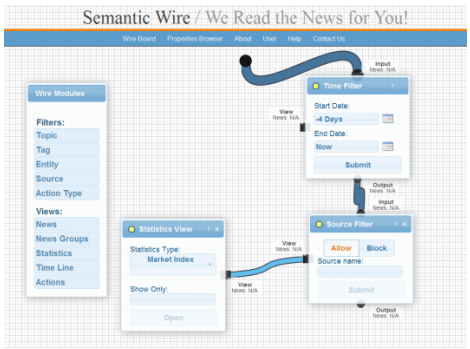


Title	**Clusty**
URL	http://www.clusty.com
Description	Clusty is a search engine with clustering functionality, based on text analysis
Functionality	entity extraction and clustering
Relevance for MULTISENSOR	• The clustering of entities could be an interesting aspect for MULTISENSOR
Similar tools/sites	• http://www.cluster-text.com (clusters whole documents according to topics)



Web News Filter

This category of tools allows for individual news channels by customizable filters that can be connected according to the user's requirements.

Title	**Semantic Wire**	
URL	http://www.semanticwire.com/	
Description	Semantic Wire is a web news analysis tool based on the Open Calais framework. Analog to yahoo pipes it offers the possibility to combine different analysis modules to get an individual filtering and output format.	
Functionality	Filtering of web news from different sources with different analysis steps	
Relevance for MULTISENSOR	• The framework • the modular structure • the visual interface	
Similar tools/sites		

Summarisation and Automatic Conversion

Title	**Gui.de**	
URL	http://gui.de/	
Description	Gui.de is an online tool that automatically creates videos from text-bases content text-based content	
Functionality	Text-analysis, automated video creation	
Relevance for MULTISENSOR	• Automatic Summarisation from text	
Similar tools/sites		

Interview questionnaire

The following Questionnaire was used for the interviews with media & journalism experts.

MULTISENSOR Questionnaire

The Internet, Social Networks and the increasing number of open accessible databases offer a fundamentally new availability of information and data. The sheer amount, the complexity, its heterogeneous state, different languages and idioms however make it nearly impossible for journalists to identify process, understand, analyze and relate all this information manually and in real time.

This is where the MULTISENSOR is trying to bridge a gap in user support. In order to align our research with real world requirements we would like to ask you a couple of questions regarding our main areas of work.

1) MULTISENSOR will start out with a search, allowing you to go through online articles, videos & audio as well as social media entries. How should such a search look like and what elements are a must-have for you?

A. A simple keyword search like Google?

B. A Keyword search with additional options?
If yes, which ones would you need?

- ☐ *Sources to search (e.g. only online-newspapers, only social media)*
- ☐ *Search period (start- & end-date)*
- ☐ *Languages*
- ☐ *Location (which geographic area you want to search in)*
- ☐ *Search for only articles, comments or both?*
- ☐ *Choice of language vs. just English or German?*
- ☐ *Other options:* _____

2) What do you think on further options to filter/refine the search results?

A. Refinement/Filtering of search results by

- ☐ *time*
- ☐ *places (only places available in the results)*
- ☐ *people*
- ☐ *events*
- ☐ *organisations*
- ☐ *languages*
- ☐ *Location (as mentioned in question 1)*
- ☐ *sources*
- ☐ *Other options:* _____

B. Possibility to reset one or more criteria from your search?
C. Possibility to save a search pattern for later reuse?
D. Possibility to combine several saved search patterns into one?
E. How would you expect the filtering to work?

☐*As a refinement of the existing list of results?*
☐*As a completely new search query?*
☐*Other options:* _____

F. Are there further steps (like refinement), that you would consider useful/necessary?

3) Sentiment Analysis allows the categorization of information (articles, tweets, etc.) into positive, neutral and negative groups. This can be done on the basis of words, sentences or whole articles.

A. Is this categorization from your point of view useful for journalistic work?
B. What other formats of sentiment analysis can be useful?
C. Assuming we consider it a useful analysis method, what would help you most?

☐*The intensity (Sentimentality) of the sentiment of information*
☐*The polarity of the sentiment?*
☐*both*

D. On what level should this analysis be conducted?

☐*In relation the article, etc. as a whole?*
☐*In relation to paragraphs?*
☐*In relation to sentences and phrases?*
☐*In relation to single objects in the article (people/places, institutions …)?*

E. How useful is it to highlight all sentiment findings (all positiv, negativ and neutral markers)? Would you rather have only the overall view of the sentiment?
F. In which Journalistic Scenario could a Sentiment Analysis be useful?

4) To assist with and speed up the manual processing of search results, especially getting a quick overview of long articles, MULTISENSOR is planning to offer automated summarizations in the form of abstracts.

A. The summaries will be done automatically, including machine interpretation.

☐*Do you consider this kind of support from the system as useful and trustworthy?*
☐*If yes, at which point would this be useful to you in the research process?*

B. Does it make sense to summaries several texts, e.g. a selection covering a similar topic, in one abstract?

B. Does it make sense to summaries several texts, e.g. a selection covering a similar topic, in one abstract?

C. What format of summary do you consider most helpful (in terms of articulation, ease-of-read, completeness)?
→ concrete example for summarization

D. How important is the connection between the summary and the original articles?

☐*Very important.*
☐*Not important.*

E. What kind of connection would you expect and what would be most useful?

☐*Simple link leading tot he full article*
☐*Linked keywords, leading to the exact position in the article*
☐*both*
☐*Further options:* _____

5) After having performed a search process the system presents an overview of the information found according to your criteria (articles, comments, tweets, etc.). What additional details could be useful in order to grasp the content in its entirety more quickly?

A. Displaying details extracted from the search results such as:

☐*All people mentioned in the results*
☐*All places mentioned in the results*
☐*All organizations mentioned in the results (e.g. Ministries, Companies, etc.)*
☐*All events mentioned in the results (Events, Conferences, etc.)*
☐*Other suggestions*_____

B. Which additional information about these details would be helpful?

☐*Relations between these entities, mentioned under A)?*
☐*Sorting of the entities for example per occurrence?*
☐*Highlighting of the entities in the results?*
☐*Links between the entities and the search results?*
☐*Other suggestions* _____

C. How could a visual representation of these entities look like?

☐*Lists sorted alphabetically*
☐*Lists sorted by occurrence*
☐*Tag-clouds*
☐*Other formats* _____

D. As an additional functionality it is possible to combine these entities with other results or analytics-methods (e.g. sentiment, summaries).

☐*Which combinations do you consider useful?*

6) Which ones of the following Metadata do you consider relevant for your search results?

☐*URL*
☐*Title of the website/blog/medium?*
☐*The authors name?*
☐*Background info on the author?*
☐*Time and date of publication?*
☐*Geographical origin of the article/tweet?*
☐*Original language the article was written in?*
☐*Different versions of the article (if existing)?*
☐*Key aspects/categories of the article?*
☐*Other options* _____

7) When visiting modern websites you often find suggestions to other relevant content, much like the Amazon-idea.

A. How useful is this concept for the journalistic research process?
B. Which information would be useful?

☐*Background information about a found entity (e.g. biography of a person, details on an event or an organization)*
☐*Comments from social media?*
☐*Other options* _____

C. Which format would you consider relevant?

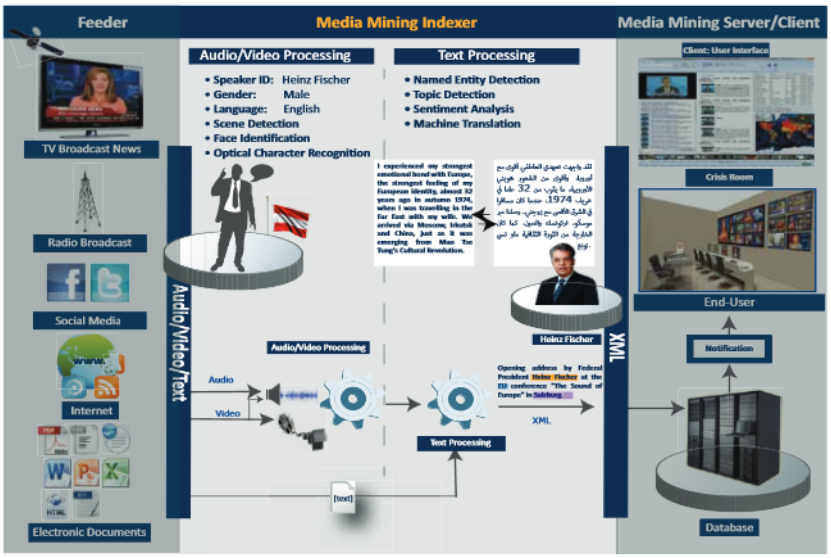
☐*Related video material?*
☐*Related audio material?*
☐*Related articles?*
☐*Other options:* _____

Thank you for your time and effort!
Are you interested in being put on our mailing list?

A.1.2 Pilot Use case 1: Media monitoring

Market analysis

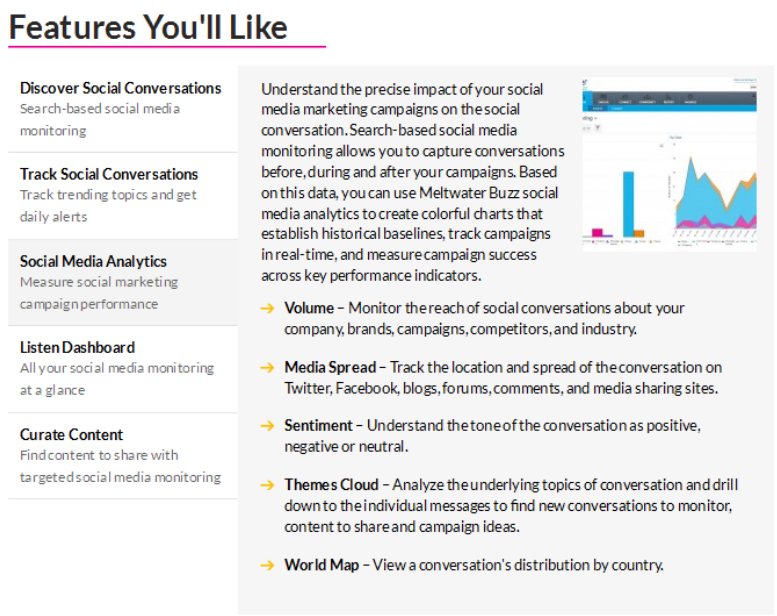
Multilingual Technologies

Title	**SAIL LABS**
URL	http://www.sail-labs.com
Description	**SAIL LABS** ("Speech-Artificial-Intelligence-Language-Laboratories") is one of the world's leading innovators in the field of **speech technology**, creating high-end software for speech and multimedia analysis solutions. The Media Mining Indexer is based upon SAIL LABS' own speech recognition technology and provides a suite of best-of-breed components for multimedia processing, transforming audio and video data into indexed and searchable information in real-time.
Functionality	**SAIL LABS** real-time multilingual automatic speech recognition technology (Automatic Speech Recognition (ASR)) • Keyword Translation, supported languages: Arabic, Bahasa Indonesia, English (US/UK), Farsi, French, German, Greek, Hebrew, Italian, Mandarin Chinese, Norwegian, Polish, Romanian, Russian, Spanish • Automatic Speaker ID / Clustering: The Speaker Identification (SID) system identifies speakers or the speakers' gender • Machine Translation; Language Model Toolkit (LMT): build own vocabulary • Data can be automatically indexed and enriched e.g. Named Entities Detection, Topic Classification and categorisation, Linguistic origin of speaker for TEXT MINING INDEXER Data Cleaning (the relevant portions of a text document are extracted and the text is standardized e.g. expanding abbreviations), Story Segmentation (the text output is segmented into coherent stories), Topic Detection and Sentiment Analysis (positive, neutral, negative)
Screenshot	
Relevance for MULTISENSOR	Speech-to-text-conversion, automatic multilingual transcription and machine translation, speakers identification, automatic analyzing data
Similar tools/sites	**Meltwater's** software understands the message and sentiment of articles from 17 different languages: http://www.meltwater.com/products/meltwater-news/online-media-monitoring/ **TVEyes** real-time transcription and translation from multiple languages into English and ability to easily translate between multiple languages. (supported languages: UK English, US English, German, Spanish, Italian, French, Chinese, Greek, Turkish Russian and Arabic) http://www.tveyes.com/?page_id=102

Sentiment and context analysis

Title	**iQ Media Group**
URL	http://www.iqmediacorp.com/pdf/iQNumbers.pdf
Description	An algorithm for detecting sentiment, using the words surrounding the search term
Functionality	• An algorithm based on the quality of language surrounding the search term and the determination of the sentimentality of the words that are in proximity to the result. • Rather than allocating a three-bucket approach (positive, neutral, negative) a score approach was created. The scores can be customized for clients based on their needs. The score allows us to differentiate the determination of positive or neutral, negative or neutral. Some clients call positive sentiment as anything that is not negative and vice versa. We can set the score cut offs based on client needs. We can also pick up client specific terms or phrases that are specific to their industries.
Screenshot	
Relevance for MULTISENSOR	• Automatic Sentiment Analysis
Similar tools/sites	

Title	**Meltwater**
URL	http://www.meltwater.com
Description	**Meltwater** provides online intelligence solutions Meltwater NEWS (Online), Meltwater Buzz (Social Media), **software-as-a-service "Just set it and forget it"**

Functionality	**Meltwater:** Sentiment Analysis - Summary of articles by positive, negative or neutral tone, extraction of topical themes, volume and geographic data DETAILS: • **Volume** – monitor reach of social conversations about your company, brands, campaigns, competitors, and industry. • **Media Spread** – Track the location and spread of the conversation on Twitter, Facebook, blogs, forums, comments, and media sharing sites. • **Sentiment** – Understand the tone of the conversation as positive, negative or neutral. • **Themes Cloud** – Analyze the underlying topics of conversation and drill down to the individual messages to find new conversations to monitor, content to share and campaign ideas. • **Word Cloud** - Visualisation of topical themes in recent press • **World Map** – View a conversation's distribution by country. • **Email Reports and Daily Alerts**
Screenshot	
Relevance for MULTISENSOR	• Automatic sentiment analysis
Similar tools/sites	• **Business Wire Inc**. distribution of press releases and **per-press-release reporting** (ONLY for English-language press releases)on social media volume, sentiment, major influencers, geographic location and other data **http://www.businesswire.com/portal/site/home/social-media-reports/**

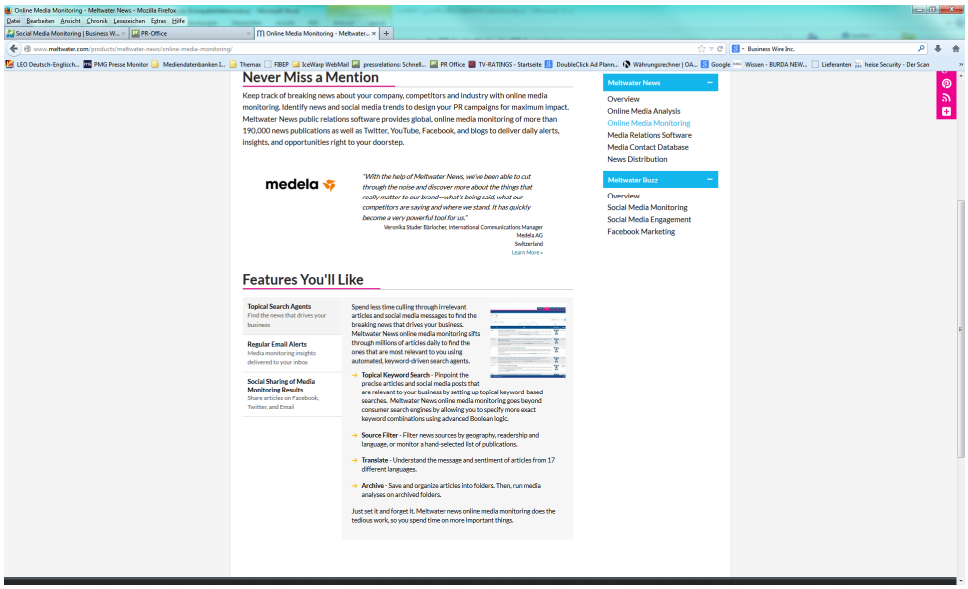
Summarise information Technologies

Title	**Newbase**
URL	**http://www.newbase.de/index.php**
Description	Customized in house monitoring systems for companies, public institutions, organisations and governments.

Functionality	**NEWBASE WebShot:** Quick and easy production of online articles, Disruptive elements such as advertising and menus are easily removed, Automatic entry of metadata, Automatic saving and entry of source names or other metadata, Direct export with all metadata into a NEWBASE system. **NEWBASE WebReview:** Search for and find your reports in the usual way using any search engine such as Google-News, Clip the selected article "on the fly" - directly in the browser, delete all disruptive elements (banners, backgrounds, menus etc.), format the clipped article, apply all the necessary metadata to the article obtained, categorize and save without the need for further software. Professional presence analysis with NEWBASE ClippingReport Service: Presence of the search term (topic, name etc.) over time, Sources and topicality, Frequent titles and sources, Word clusters in titles and text.
Screenshot	
Relevance for MULTISENSOR	• Fully automatic media monitoring (complete solution provider for media monitoring software, digitisation and conversion services and app development)
Similar tools/sites	

Crawling infrastructure

Title	**NEWBASE**
URL	**http://www.newbase.de/index.php**
Description	see above
Functionality	**NEWBASE iSpider:** • Fully automatic observing of defined websites. The program analyses websites at preset intervals and identifies new articles. • Fully automatic capture of WebPages. If a new article is discovered, the program saves the page including metadata (name of the website, topic, author etc.). • Fully automatic article clipping. In the enhanced version of the program the article is clipped from the webpage. Whereby all irrelevant elements such as advertising, menus etc. are removed automatically. Metadata (name of the website, topic, author etc.) are also saved. • Fully automatic saving and classification. The WebPages or articles are saved to the database. If you maintain reader profiles, the program automatically searches for key words in the contents. The articles are allocated to the topics on the basis of the key words found. • Release in the portal. The Internet articles can be presented in the NEWBASE Media Portal along with other articles from various media channels. In other words, all articles are available in one Intranet.
Screenshot	
Relevance for MULTISENSOR	• Fully automatic media monitoring (complete solution provider for media monitoring software, digitisation and conversion services and app development)
Similar tools/sites	

Title	**Meltwater**
URL	http://www.meltwater.com
Description	see above
Functionality	**Social Media:** Use keywords to identify specific topics of conversation that are important to your social community, such as brands names, competitors, industry phrases, acronyms, social hash-tags, celebrities, trending topics, etc. The Meltwater Buzz Listen module's search-based social media monitoring lets you dig deep into the social media chaos to target the most relevant conversations using **precise Boolean logic**. **Online:** Meltwater News online media monitoring goes beyond consumer search engines by allowing you to specify more exact keyword combinations using **advanced Boolean logic**.
Screenshot	
Relevance for MULTISENSOR	• Finding the most relevant content using automated, keyword-driven search agents.
Similar tools/sites	

Graphic interfaces

Title	iQ Media
URL	• http://www.iqmediacorp.com/pdf/PL2-Media-Optimized-for-the-Cloud.pdf • Demo • cliQ 4.0 • iQ Media YouTube channel
Description	• Industry's first Media Intelligence Platform, enabling users to leverage TV video, online news, and social media content optimised for the cloud. • Users do not need to invest in IT infrastructure or manage the network. They have instant access to media at all times from any Internet- capable platform. • "Big data" is the key to unlocking the content marketing potential of that media. We developed cliQ with the notion that the more data we can ingest, organize, and make available, the more our customers can do with it. • Software-as-a-Service solution: cliQ's cloud-based media is based on a software breakthrough developed by iQ media, the Lakshmi Scalability Algorithm (LSA).
Functionality	• cliQ: Cloud based -> Media is optimised for availability, accessibility, and quality. cliQ makes media available far longer, accessible far more quickly, and presented in a higher quality format. • Total Recall: Media is available for longer periods — even indefinitely. Dated TV coverage and online news that suddenly becomes important can still be accessed months later in the cloud. • Fast Federated Searches: Content is pre-optimised for rapid access across all media type. Users get results quickly, pinpoint what they need, and repurpose media for sharing and permanent storage. • Simple Sharing: Users can quickly email links to video clips and other media, post them to social media platforms, or stream them from their own websites. • Extended Management: Users can put their own user-generated video content into the cloud, where cliQ will apply the LSA to optimize it.
Screenshot	
Relevance for MULTISENSOR	• Blueprint for platform setup
Similar tools/sites	• Meltwater

Interviews

Interview 1: MediaImage, Romania

Interviewee(s): Miron Mateescu, Vali Purecel, Sorin Milutinovici

As media monitoring company, what are the requests you are mostly confronted with?

As media monitoring company for the Romanian, Moldavian and Bulgarian market, MediaImage is expected to track all relevant online news in this area. The clients mostly ask for quantitative tracking of news. A qualitative evaluation or a qualitative press review is requested only in very few cases. Thus, the main task is to monitor all online media.

What are the most time-consuming factors in your daily work?

The delivery of press clippings is organised in a decentralised way: MediaImage has employees in several cities in order to get hold of regional dailies. These media are read; articles are scanned and subsequently delivered to Bucharest, where they are processed further. This process is very time-consuming, and therefore the complete press review (consisting of web and press articles) can be delivered to the client only relatively late in the day.

What metrics are of special importance to your clients?

In order to be able to correctly assess which online media will potentially release a large number of relevant news for a given client, the overall number of news released per month/week by the media is important.

In your opinion, what are the greatest challenges in handling social media?

According to MediaImage, social media is requested seldom if at all. As media monitoring company, they sometimes deliver Twitter or Facebook content as an add-on to the web search. These results are received as "nice to have" by their clients, who mostly lack to knowledge how to handle this content. In this regard, the biggest challenge lies in promoting the social web as a relevant media source to the client and in showing them the additional value and evaluation possibilities.

What kind of content Visualisation would be of help when performing your daily work?

There is a problem with Visualising clusters with identical articles. Most important is that it should be possible to see at first glance how many articles from which media sources are contained in a cluster.

What kind of product would be truly innovative from your point of view?

MediaImage's business model is based on a price model with flat rates. This is opposed to a relatively big amount of articles that need to be processed and delivered. For this reason, automated processes are hugely important: automated capture, automated coding, automated clustering, and automated sentiment analysis.

Interview 2: MMO, Austria

Interviewee(s): Edith Wilflingseder

As media monitoring company, what are the requests you are mostly confronted with?

Standard media monitoring for Austria (Press, Online, Radio, TV), test monitoring free of charge, licensing issues (VÖZ, PMG, etc.), research, media monitoring conditions abroad, questions whether special trade press media are included in the media sample

What metrics are of special importance to your clients?

Diverse media data: distributed edition, release dates, release intervals, advertising rates and advertising equivalence value, article editors, magazine covers in addition to the articles themselves, the whole page of the clipping as opposed to the clipping itself.

In your opinion, what are the greatest challenges in handling social media?

Quantity vs. quality: due to the rapidly growing media resonance, the client needs to be informed as fast as possible, yet in a clear way. Alerting services supplemented with a media resonance analysis are able to fulfill these tasks, but often exceed the client's order volume. Automated tool are only of limited use for the range of individual client requirements.

What kind of content Visualisation would be of help when performing your daily work?

A Visualisation structured by key words, topics or sectors; display of identical news items only once and listing all additional sources; dynamic creation of charts and graphs for individual queries (pie, bar, and line charts etc.); an index of online and social media news in the PDF for later text searches.

What kind of product would be truly innovative from your point of view?

A tool that automatically implements individual requirements, meaning a search profile that automatically considers all restrictions and requirements after an initial setup.

What are the most time-consuming factors in your daily work?

Implementation of client search requests; calculation and implementation of special solutions; performance of international media monitoring requests.

Interview 3: PressService, Poland

Interviewee(s): Marcin Szczupak

As media monitoring company, what are the requests you are mostly confronted with?

In terms of media analysis it is mostly the measurement of efficiency of media relations activities, supporting the client with a better understanding and using of media, constant (ongoing or even preventive / in advance) informing the clients about key news about them.

What metrics are of special importance to your clients?

Unfortunately, this is still the number of clippings (publications) and the AVE. However, more and more often we discuss and negotiate with the clients on more advanced indexes, e.g. PRIME, which are based on individual agreements (findings) with the clients. What is

important is that these discussions often concern measuring the ROI in media communication (media relations) as compared to other communication activities (marketing, advertisements, sponsoring). We try to standardize the way of measuring the efficiency.

In your opinion, what are the greatest challenges in handling social media?

Providing good quality of media monitoring – finding each and every critical or unfavorable piece of news; correct selection of important news and leaving those less important. The volume of information in social media is so big that this is a crucial task.

Another task is measuring the efficiency of media communication and establishing, what is the relation between the news and the actual media image of a company / product / brand.

What kind of content visualisation would be of help when performing your daily work?

If the question is related to preventive aspect and measuring of the efficiency – there are a lot of elements, which could make our work easier. Growing media activity in particular subjects, bigger activity of users / media / journalists / subjects: signalling the trends in particular subjects.

What kind of product would be truly innovative from your point of view?

A tool which would allow to recap and analyze all media and which would allow to manage all media (in terms of communication, but also in terms of the know-how of these media, including social media).

What are the most time-consuming factors in your daily work?

Building new methodologies, creating individual solutions for the client, team management, quality control and quality measures.

Interview 4: Lautenbach Sass, Germany

Interviewee(s): Katharina Simon, Maren Müller

As media monitoring company, what are the requests you are mostly confronted with?

Clients sometimes wish to position themselves as opinion leaders for defined topics. For this reason, rather broad monitoring of topics is requested in addition to the monitoring of the company itself and its competitors. An easy and structured tracking of topics is essential.

Also, we are often confronted with requests for retrospective analysis. Tracking and analysis of past media coverage has sometimes been difficult.

What metrics are of special importance to your clients?

Clients are especially interested in opinion leaders. Sometimes also other criteria such as picture penetration and article style are considered relevant.

In your opinion, what are the greatest challenges in handling social media?

The volumes are really huge and it is important to track the actually relevant influencers, who have large communities and networks behind them. Presently there is no fast and reliable way to determine whose posts reacting to would be crucial. Also, the colloquial language can be a problem for analysis. Then, there is also the principal question, whether every single post should be considered in isolation or within the context of the whole thread.

What kind of content visualisation would be of help when performing your daily work?

Tag clouds are seen as helpful tools, especially when it is possible to navigate within them. The same goes for visualisations of networks (especially in social media). Visualisations of quotations in other articles could also be of use.

What kind of product would be truly innovative from your point of view?

It is interesting for the client to determine whether the coverage he gets can be directly traced to the client's PR or if it is actual journalistic content. A system that can trace quotations and plagiarism would be valuable, as it is hard to manually check huge amounts of content against e.g. client press releases. This would ideally be combined with a Visualisation of the dependencies. One could also imagine an interface, where the client could enter his own press releases and PR material for cross-check with the media coverage.

What are the most time-consuming factors in your daily work?

A detailed analysis takes a lot of time: when looking at significant peaks in the coverage, we usually need to reread a lot of articles to refine the insights. It would be great to have system support in form of keywords or a written synopsis for the article selection that can be further refined on demand. Another possibility would be to determine the most representative article out of the selection.

Interview 5: MTM, Switzerland

Interviewee(s): GianCarlo Bianchi, Stephan Wyss

As media monitoring company, what are the requests you are mostly confronted with?

As a Swiss media monitoring company, MTM mostly is confronted with the task to provide a universal solution for all available channels and for all three Swiss official languages (often with English in addition). This is a special challenge with clients who wish to order unedited media monitoring, as satisfying results frequently require a lot of work.

What metrics are of special importance to your clients?

No uniform trend can be discovered. In the area of printed media, several indicators are interesting, but the demands vary from client to client. As for online media, metrics are seldom asked for. Measurement according to e.g. visits has up to now been mostly irrelevant.

In your opinion, what are the greatest challenges in handling social media?

For MTM, the greatest challenge in the area of social media lies in the sensitisation of the Swiss clients, who up until now have displayed perceptive restraint. According to MTM, social media monitoring in Switzerland has not yet attained the status it has e.g. in Germany. MTM does, however, expect a gradual increase within the next two years.

What kind of content visualisation would be of help when performing your daily work?

As most of MTM's clients focus on press articles and usually only ask for unedited online monitoring, a visualisation of clusters is not that relevant. It would however greatly reduce the expenditure for editing and analysis if basic article topics could be easily discerned and visualised.

What kind of product would be truly innovative from your point of view?

The special wishes of the client are central to MTM's business model. For this reason, an especially innovative product would be a tool that captures these specific wishes exactly and enables them to mass tailor our products in an uncomplicated and fast way.

What are the most time-consuming factors in your daily work?

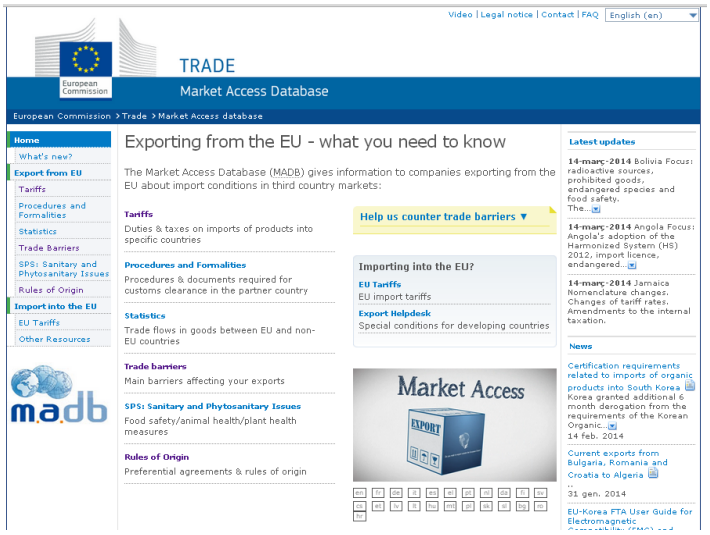
MTM perceives manual coding of articles by topics as particularly time consuming. Thus, they express a particularly strong wish for automated topic detection as coding aid, or for automated coding of online articles (as a better alternative to unedited monitoring).

A.1.3 Pilot Use case 2: SME internationalisation

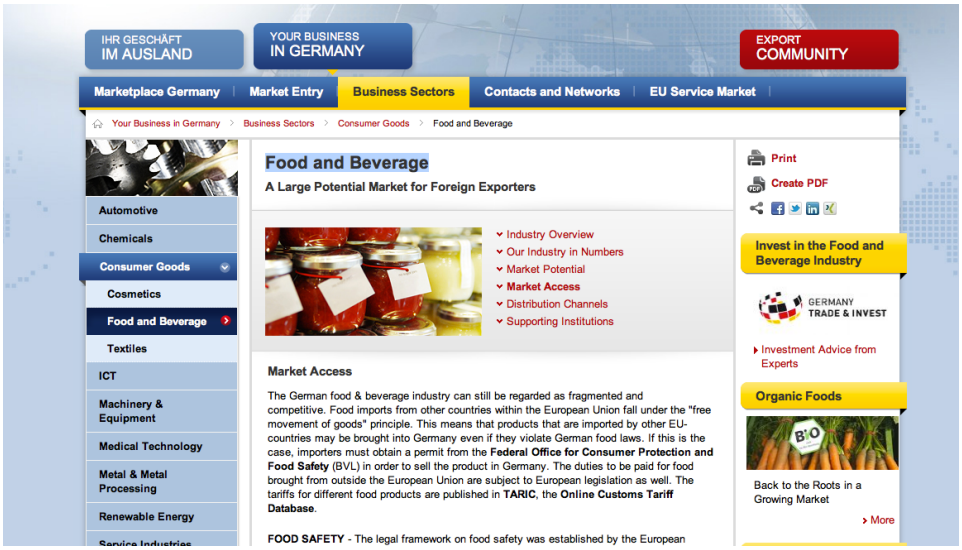
Currently there are no websites that contain the complete amount of information that the SMEs need in order to internationalise, as far as we are concerned. What we can find are websites that provide us with separated information, here below you will find some examples:

Title	**ICEX**
URL	http://www.icex.es
Description	ICEX is the website of the Spanish ministry of economy and competition. In this website they give general information about countries worldwide. Putting as an example Germany, looking at ICEX you can find General Data (about its geographic situation, climate, demography, society, brief history,…)., Political Environment, Economic Structure (which are the most popular sectors,…). There is also information about how to export to Germany (some statistics, to barriers, the best and strategic channels of distribution, some technical and juridical aspects,…)
Functionality	
Screenshot	
Relevance for MULTISENSOR	• about how to access the market, visas, city websites, bank holidays, banks,...
Similar tools/sites	• http://accio.gencat.cat/cat/ • http://www.embajada-alemania.es/ • http://www.auswaertiges-amt.de/ • https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html

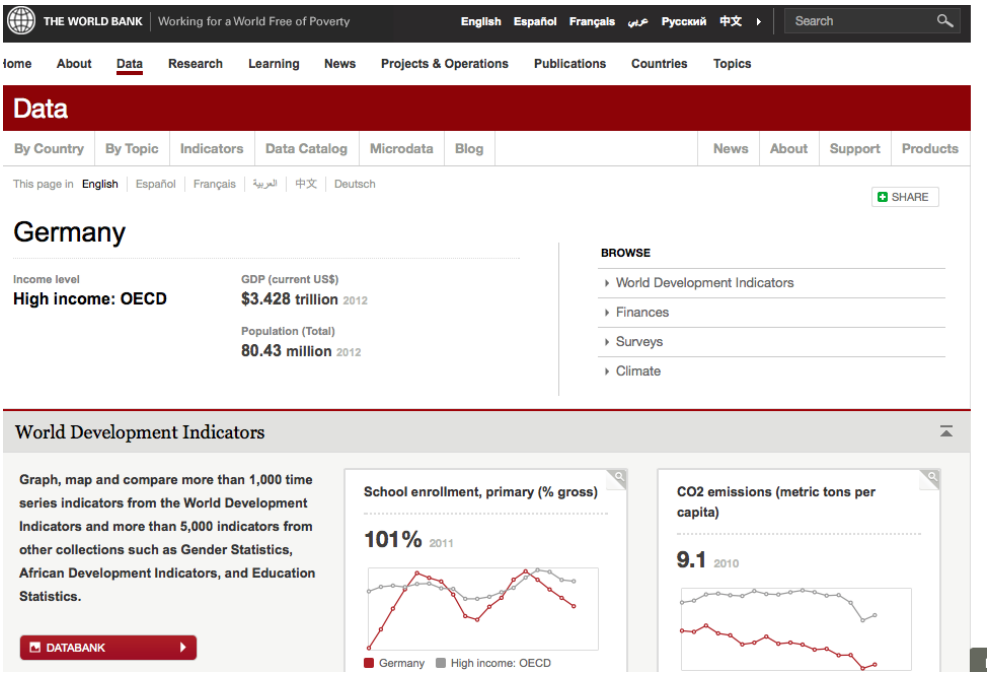
Title	**Market Access Database**
URL	http://madb.europa.eu/madb/indexPubli.htm

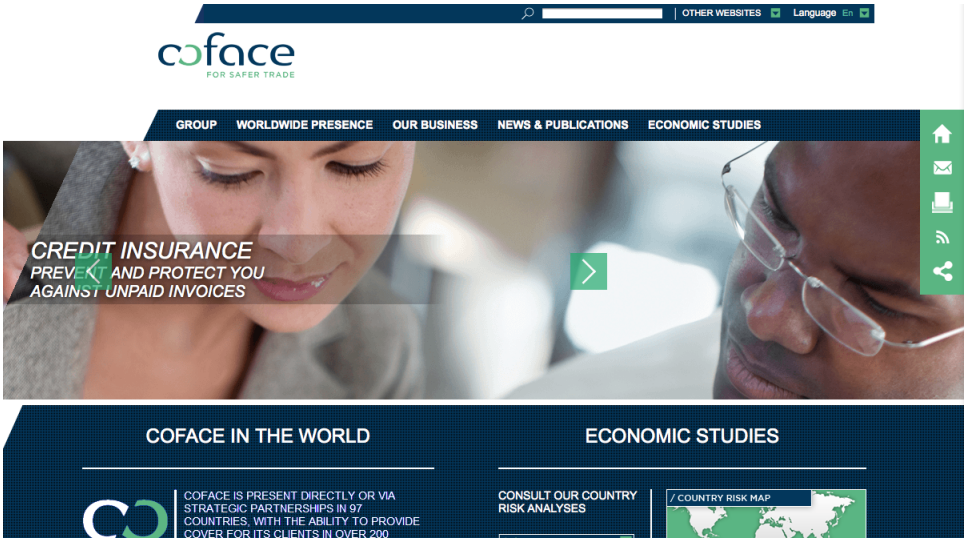
Description	Market Access Database is part of the European Commission website where it is focused on exports from the EU about import conditions in third country markets. You can find information about the tariffs, procedures and formalities, statistics, trade barriers,… This website, in comparison to ICEX, provides the possibility of specifying a little bit more since they ask for the country where you want to export and the product that you want to export. The way they find the products is by looking at a 4 or 6 digits number called Taric Number. The problem about this website is that it's only focused on third countries so there will be no information regarding Germany, for example.
Functionality	
Screenshot	
Relevance for MULTISENSOR	• documents • barriers • Taric
Similar tools/sites	

Title	**Europa, Summaries of EU legislation**
URL	http://europa.eu/legislation_summaries/consumers/product_labelling_and_packaging/l21090_en.htm
Description	Europa – summaries of EU legislation,
Functionality	
Screenshot	
Relevance for MULTISENSOR	Labelling, presentation and advertising of foodstuffs
Similar tools/sites	

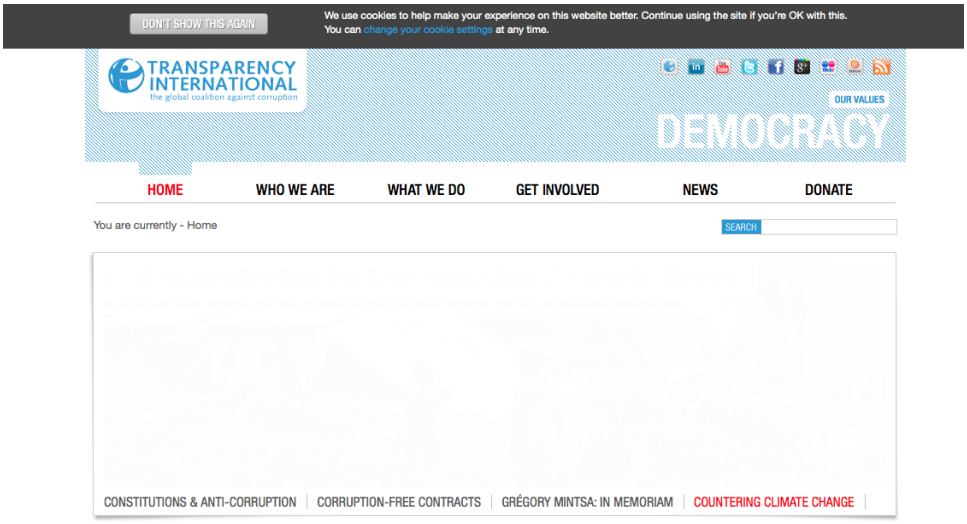
Title	**Ixpos.de**
URL	http://www.ixpos.de/IXPOS/Navigation/EN/Your-business-in-germany/Business-sectors/Consumer-goods/food-and-beverage,did=263444.html
Description	The German business portal
Functionality	• sentiment-analysis • supporting several languages
Screenshot	
Relevance for MULTISENSOR	• Information about de food and beverage market in Germany
Similar tools/sites	• http://sproutsocial.com/ • http://www.brandwatch.com/key-features/

Title	**Doing Business**
URL	http://www.doingbusiness.org
Description	The Doing Business Project provides objective measures of business regulations and their enforcement across 189 economies and selected cities at the subnational and regional level.
Functionality	Covers business regulation and reform in different cities and regions within a nation. These reports provide data on the ease of doing business, rank each location, and recommend reforms to improve performance in each of the indicator areas. Selected cities can compare their business regulations with other cities in the country or region and with the 189 conomies that Doing Business has ranked.
Screenshot	
Relevance for MULTISENSOR	• documents • barriers • regulations
Similar tools/sites	

Title	**The World Bank**
URL	http://www.worldbank.org
Description	World Bank is an international organisation to require open access to much of its research outputs under Creative Commons licensing. The information is about financial, innovation
Functionality	
Screenshot	
Relevance for MULTISENSOR	Information about countries and financial issues
Similar tools/sites	

Title	**Coface**
URL	http://www.coface.com
Description	Risk of non-payment of the countries
Functionality	To know which country has a risk of non-payment
Screenshot	
Relevance for MULTISENSOR	• Information about the risk of payments
Similar tools/sites	

Title	**Events eye**
URL	http://www.eventseye.com
Description	Trade Shows, Exhibitions, Conferences & Business Events Worldwide
Functionality	To find shows, exhibitions …..of your sector
Screenshot	
Relevance for MULTISENSOR	To be more accurate about the information of the sector
Similar tools/sites	http://www.nferias.com

Title	Transparency International
URL	http://www.transparency.org
Description	To know the Corruption of a country
Functionality	To complete the study of a country, also it is of relevancy to know his degree of corruption.
Screenshot	
Relevance for MULTISENSOR	• To complete information for the companies
Similar tools/sites	