D1.1 – Preliminary Version of Use Case Design

Document Number	D1.1
Document Title	Preliminary Version of Use Case Design
Version	1.3
Status	Final
Work Package	WP1
Deliverable Type	Report
Contractual Date of Delivery	30.04.2014
Actual Date of Delivery	30.04.2014
Responsible Unit	UNITN
Keyword List	Use_cases, user_tasks, targeted_users, topical_analysis, sentiment_analysis, authority_analysis, summarization, synopsis
Dissemination level	PU
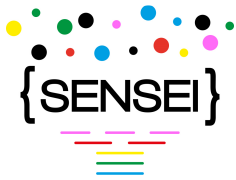
Editor

Morena Danieli   (UNITN)

Contributors

Giuseppe Riccardi   (UNITN)

Emma Barker   (USFD)

Jonathan Foster   (USFD)

Adam Funk   (USFD)

Rob Gaizauskas   (USFD)

Mark Hepple   (USFD)

Emina Kurtic   (USFD)

Massimo Poesio   (UESSEX)

Letizia Molinari   (TP)

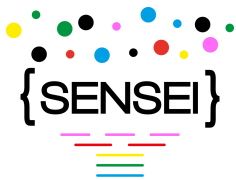Vincenzo Giliberti   (TP)

SENSEI Coordinator

Prof. Giuseppe Riccardi

Department of Information Engineering and Computer Science

University of Trento, Italy
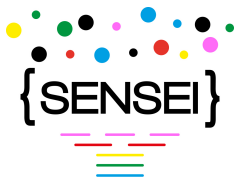
riccardi@disi.unitn.it

Document change record

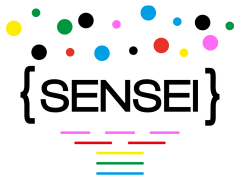| Version	Date	Status	Author (Unit)	Description
0.1	01/31/2014	Draft	M. Danieli, G. Riccardi (UNITN), L. Molinari, V. Giliberti (TP)	Table of Content, draft of the speech use case
0.2	03/19/2014	Draft	M.Danieli (UNITN), E.Barker, J.Foster, A. Funk, R.Gaizauskas, and M.Hepple (USFD)	Speech Use Case revised Social Media Use Case first draft added
0.3	04/04/2014	Draft	M. Danieli, G. Riccardi (UNITN)	Speech Use Case: final draft; added use scenarios in Section 4; editing of social media scenarios in Section 4 (added title styles) Revised Executive Summary Revised List of Contributors Revised section 4 (data gathering)
0.4	04/04/2014	Draft	Elisa Chiarani (UNITN)	Quality Check
0.5	10/04/2014	Draft	E. Barker, J. Foster, A. Funk, R. Gaizauskas, M. Hepple, E. Kurtic (USFD)	Social Media Use Case completed
0.6	11/04/2014	Draft	E. Chiarani (UNITN)	Second Quality Check
0.7	11/04/2014	Draft	M. Danieli (UNITN)	Consolidated draft ready for scientific review
0.8	13/04/2014	Draft	M. Danieli (UNITN)	Added draft of Section 3
0.9	14/04/2014	Draft	M. Poesio (UESSEX)	Scientific Review
0.10	15/04/2014	Draft	R. Gaizauskas (USFD)	Minor corrections to Summary and Section1 and revisions to Section 3
0.11	16/04/2014	Draft	G. Riccardi (UNITN)	Review
1.0	23/04/2014	Draft	M. Danieli (UNITN)	Updated and consolidated version after review
1.1	27/04/2014	Draft	M.Danieli (UNITN)	References added to the speech use case section
1.2	29/04/2014	Draft	A. Funk (USFD), E. Chiarani (UNITN)	Final quality check and review
1.3	30/04/2014	Final	M.Danieli (UNITN)	Keyword list added. Definitive version.

INDEX

Executive summary

In this deliverable we present the preliminary version of the use cases for the social media and speech domains. This document provides the complete scope of the scenarios that will be considered in the SENSEI project. In fact the follow-up deliverable D1.2 will provide a selection and revision of the scenarios from this report to be evaluated. The deliverable is designed to inform the other WPs of the project with information, descriptions, and goals that are essential for focusing their data collection, annotation and technology design efforts. The goals of D1.1 are as follows:

1. to identify target end-users of the Sensei prototype versions;
2. to illustrate a set of use cases for the social media and speech domains;
3. to provide input to the rest of the project work packages for finalizing the descriptions of the use cases at the end of Y1.

The document has three sections: one for each domain of application and a third one discussing similarities and differences between the two application scenarios. The first section is devoted to a discussion of the issues related with the speech scenario; the second section is devoted to the news and social media scenario. Each of these sections identifies user groups, provides a preliminary description of their needs, proposes candidate document summarization and analytics required functionalities, and illustrates further work to be done in the second part of Y1 for gathering user requirements.

1. Speech Use Case Introduction

Consumer-oriented companies usually delegate their customer touch-point operations to outsourcing call centre companies. Call centres manage huge amounts of information whose processing can result in data that is critically relevant for business, and that would otherwise be lost. By analysing and categorizing data extracted from on-line listening or from the recorded interactions that customers have with the call centre, call centre companies and their clients can get insights related to marketing strategies, products, processes, operations, and call centre agent issues. The corporate clients of call centres may also require reporting in different aggregated forms, according to, among others, the topic of the calls and the emotional attitude of callers with respect to them (e.g. Bailor 2006).

In call centres, in-bound and out-bound calls are either monitored in real time or recorded for later review. Both these activities are carried out by human quality evaluators, who can only check very small, randomly selected samples. The job of these evaluators is to monitor the quality of agents' work and to track global indicators of call quality and efficiency. They fill out forms that include behavioural and conversational indicators, such as politeness, listening attitudes, empathy, language use and ability to keep the interaction focused.

In outbound marketing campaigns, the data collected by human evaluators are used for assessing not only the performance of the call centre agents, but also for getting early evaluation of the results of the marketing campaign.

In this section we describe categories of potential users of the SENSEI speech scenario, the information needs of these potential users, candidate use cases, methods for gathering feedback on the preliminary use cases and functionalities that the technologies developed in the project can provide to potential users.

1.1 Categories of Users

Two main categories of potential users of SENSEI technologies for inbound and outbound call centre activities have been identified:

- **Strategic management - business intelligence teams:**

  1. Quality & Processes Area;
  2. Accounting Area;
  3. Quality Assurance Management;
  4. Human Resources Area.

- **Operations management:**

  1. Management Front Office;
  2. Quality Assurance Management;
  3. Quality Assurance Supervisor;
  4. Training Management;

Within the first main category, Strategic Management, we have identified four subcategories of professionals:

- the professional who needs to listen to calls for setting standards and guidelines for call centre agents;

- the professional who needs to listen to calls for setting standards and guidelines based on client requirements;
- the professional who needs to listen to calls for evaluating their overall quality and compliance to guidelines (front office supervisors);
- the professional who needs to listen to calls to identify needs and requirements relevant to the professional growth of employees.

Within the second main category, Operations Management, we have identified the following subcategories of professionals:

- the front office manager, who needs to listen to calls and share output from Quality Assurance (QA henceforth) to ensure global guidelines and client requirements are met;
- the QA manager who needs to listen to calls for sharing output from QA to ensure the respect of global guidelines, of client requirements and calibration between QA evaluators;
- the QA evaluator who needs to listen to calls and measure agent by agent the global level of management of the calls;
- the Training manager who needs to listen to calls and receive output in the form of a global report in order to plan specific training activities.

The *standards and related guidelines* mentioned above are usually set by the QA department. They apply to call centre agents, and include how the call is answered, how well the agent engaged the customer, how well the agent used available resources, whether or not the customer's issue was resolved and how the agent closed the call. The QA department usually puts together a telephone script for agents to follow on each call. It is worth noticing that a particular 'rapport-building' speech style is often prescribed to call centre workers across different countries, irrespective of the language being spoken in the language interaction (e.g. Hulgren 2011).
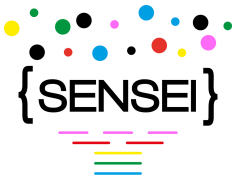
1.2 Listening to act and listening to learn

The descriptions provided in the previous paragraph show that while listening to the calls constitutes the basis for most of the tasks of call centre professionals both in the Operations and Strategic areas, yet those inspection activities can be carried out by assuming several different observation vertices, mostly depending on the objectives of the professional tasks. The structure and content of each randomly chosen conversation during listening activities is likely to be considered in detail in order to identify topic, client and agent behaviour, task success, and compliance with QA guidelines and scripts.

If we consider the call centre professional activities from the point of view of their kind of listening, we can try to abstract from those subtle differences and possible duplications. For example, we can assume that call centre conversations can be listened to for getting pieces of information useful for acting (listening to act). An example of this kind of listening is the listening done for setting guidelines and requirements. Listening to act is performed by professionals of both main areas we have identified, i.e. Strategic Management and Operation Managements, with different goals. The professionals in the first area have the goal of setting, or adjusting, standards and guidelines (activities 1-3 in the above relevant list), while the professionals in the second area, i.e. Operations, have the goal of evaluating call centre activities, operationalizing guidelines, and evaluating agent compliance to guidelines.

In both main areas the fourth category of professionals are likely to perform the kind of listening that we call listening to learn. Actually the goal of those professionals is to capture information for identifying training needs, critical agent behaviours, coping strategies, etc., in order to define and plan strategic and operational education interventions (e.g. Liao et al. 2009).

In the following we will leverage this distinction for identifying examples of user requirements for the different user categories.

1.2.1 Listening to act for setting and evaluating call centre quality standards

QA professionals, in both the Operations and Strategic Management areas, are responsible for evaluating the quality standards for incoming and outgoing calls. The QA team spends a large amount of time each day listening to live calls. While listening to a call, the QA professional refers to checklists to determine if the agent handles the call according to QA standards and guidelines. Based on the results of the checklists, the agent is assigned an overall score for the call and is then notified of his score. This information can also be used for identifying training needs for groups of call centre agents, and for focusing the educational contents on such needs

The primary task of business intelligence reporters is to gather the information required for monitoring the key performance indicators (KPI) used to assess customer satisfaction. Each call centre should abide by basic performance indicators like customer satisfaction or lack of satisfaction, call topic, time saved, and quality. For example, in addition to time spent by each agent in calls, other indicators may give better understanding of how the call centre performs, for example high rate of unanswered calls due to increased volume of calls will result in frustrating the customers, poor audio quality and repeated questions may lead to calls being abandoned.

In addition to traditional KPI such as rate of abandoned calls, time spent by the agents on call, and silences in conversation, these professionals may be interested in any indicators useful to assess directly customer emotional attitudes, such as frustration, satisfaction, and how the customer emotions can be affected by the empathy of the call centre agents. Actually, despite a long period of neglect, research on emotion in organizational behaviour has developed into a major field over the past recent years, and is now seen to be part of an 'affective revolution' in the organization sciences (e.g. Ashkanasy 2011; Jenkins et al. 2010). For the SENSEI targeted users the evidences about emotions need to be related to information about the semantic content of the calls, including the call topic (e.g. Baldoni et al. 2012; Alam & Riccardi 2013).

1.3 Sensei Use Cases for the Speech Scenario

1.3.1 Use Case 1: Assistance while monitoring agent-customer conversations

Actor: Supervisor of the Front Office

Goal: to get support while evaluating the overall quality and compliance to guidelines of on-going calls

Steps:

1. The Supervisor is listening to a conversation going on between one of the supervised agents (randomly chosen) and a customer. Usually the Supervisor of the Front Office needs to continuously check the quality of 8 to 10 call centre agents.

2. Supervisor indicates to the system a critical aspect of the conversation (conversation areas) s/he needs assistance with, such as opening, call management, closure, and time.

3. The system analyses the conversation area(s) of user interest in the on-going calls, based on different indicators for each area, and prompts the Supervisor when one or more on-going conversations report critical values for the selected area. For example, if the Supervisor selected "Call Opening" or "Call Closure" as the conversation area of interest, the system checks if the Agents are following the scripts provided them for opening and closing. If the Supervisor selects "Time" the system monitors the call time and provides feedbacks related to the calls that report call time values different from the expected average values. If the Supervisor selects "Call management" s/he will need to provide the system with further indications about the aspects to be monitored. For example, if the Supervisor needs to monitor the fact that the agents are trying to sell new commercial products while managing inbound calls, the system will provide feedback from time to time with a brief report containing the number of commercial offers occurring in the on-going calls.

1.3.2 Use Case 2: On-line monitoring of sentiment trends in the conversations

Actor: Front Office and QA Supervisor

Goal: to determine emotional trends in the on-going conversation

Steps:

1. The Front Office Supervisor is listening to a conversation going on between one of the supervised agents (randomly chosen) and a customer; the supervised agents are usually 8-10 for each Front Office Supervisor.

2. The supervisor indicates to the system a critical emotion, such as Empathy, Frustration, Anger, s/he needs assistance with. S/he also indicates if the monitoring applies to the Agent channel or to the Customer one.

3. The system analyses the emotion of user interest in the on-going calls, and prompts the Supervisor with feedback from time to time when one or more on-going conversations report critical values for the selected emotion.

1.3.3 Use Case 3: Reporting for the Quality Assurance Managers and Professionals – Conversation oriented summaries

Actor: QA Professional and/or QA Manager

Goal: Getting daily feedback in the form of reports including conversation summaries oriented to extract indicators concerning agent behaviour with respect to their compliance to the QA scripts.

Steps:

1. The QA professionals identify a call centre campaign they will need assistance with, and provide the system with the QA scripts that needs to be observed by the agent in the campaign. The scripts include opening and closure specifications, emotions to be monitored, semantic content such as commercial offers, etc.

2. The system performs analytics on recorded data with respect to the specific aspects identified by the QA professionals; reports present data either in graphical form or in the form of summaries that report conversation contents related to the aspects of interest and refer to aggregations of calls based on such semantic contents.

1.3.4 Use Case 4: Automatic generated call surveys

Actor: QA Professionals and/or QA Manager

Goal: Increasing the number of monitored agent-customer calls

Steps:

1. The QA professional can only listen to a very limited set of inbound and outbound calls out of the thousands going on in an outsourced call centre each day. S/he may profit from automatically generated surveys for a greater number of monitored calls.

2. The manually generated surveys currently used by QA professionals are organized around several areas of interest, including Call Opening, Problem Identification and Fixing, Commercial Offers, and Communication Skills.

3. Each area of the survey refers to some questions, whose answers may be Yes, No, or N/A.

4. While replying to many of those questions implies sophisticated human capability to evaluate agent behaviour, nevertheless the job of QA professionals could benefit from the automatic generation of KPIs, such as the degree of script compliance by the agent, the management of waiting time, the occurrence of speech overlap during the call, the classification of the calls according to the possible presence of positive or negative emotions.

1.4 Gathering feedback on the preliminary speech use cases

The preliminary use cases for the speech scenario described in paragraph 1.3 need to be evaluated, enriched, and developed on the basis of assessment of their sustainability, and on the basis of user acceptance. The first evaluation is "project internal": the proposed preliminary version of the use cases for the speech scenarios will be revised and discussed within the SENSEI Consortium with a view to prioritizing the implementation of the most promising ones.

In parallel, within WP1 for the speech scenario we will gather data and feedback from the potential users identified in section 1.1. In particular, we will interview QA and Operation managers we have already contacted among the Teleperformance staff. We will define a loosely structured set of questions that will be addressed to the interviewed persons either in person or remotely.

We also propose to assemble focus groups. For these we plan to recruit different sets of subjects including front office QA professionals, QA and operation managers of call centres. In these focus groups we will ask participants for open-ended responses conveying their thoughts or feelings about the way they view the potential benefits from applying analytics in their daily activities. The focus group moderator will pose questions in a way that can stimulate insightful answers. From qualitative social research methodology we know that focus groups can be useful when the research aims to stimulate comments that can emerge from free-flowing confrontations of different views. We think this will be useful for fixing user requirements and desired functionalities.

1.5 Sensei functionalities for the speech scenario

1.5.1 Conversation oriented summaries (synopsis)

Based on clusters of conversations, *conversation-oriented summaries* (*synopsis*) will be able to provide the different categories of potential users with a range of indicators that are useful for the different listening processes illustrated above. They can support listening to act by providing
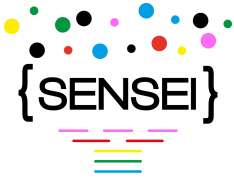
summaries that will allow conversation clusters to be evaluated on the basis of pragmatic and para-semantic features, and of emotional polarity of the groups of conversations. In addition they will allow conversation clusters to be related to different variables, such as call topic and operator efficiency.

We expect that conversation oriented summaries may enable the discovery of unknown correlations across different observed variables. Specific sets of queries will be represented in terms of data aggregation and selection over datasets of annotated calls.

1.5.2 Reports and rated questionnaires

The listening to act activities of call centre professionals can also be enhanced by providing them with *user defined ad hoc reports* that summarize the results of user supplied queries. At present, the control of call quality is based on human listening to call samples and filling in questionnaires that aim to assess the behaviour and professional competence of call centre agents. Politeness, linguistic appropriateness, clarity of expression, affective attitude, and other behavioural indicators are addressed in such questionnaires. This kind of quality assessment questionnaire implicitly includes a rating model that the professional evaluators apply to each agent's calls. Not all the calls are evaluated, but often professional evaluators need to listen to a considerable number of arbitrarily chosen calls. In SENSEI we can operationalize the implicit rating model represented in the questionnaire and will use it to score conversations of single agents or groups of agents. The professional evaluator task will then rely on the summaries obtained by these *rated questionnaires*. The expected benefits are both in terms of productivity (less human listening) and in terms of a more empirically oriented assessment process (more focused human listening).

Ad hoc reports and rated questionnaires can be generated and personalized also for supporting the listening to learn of human resources professionals in both Strategic, in particular QA, and Operations areas.

2. Social media use case

2.1 Social Media Use Case Introduction

On a news publisher website such as *The Independent* or *Le Monde*, journalists publish articles on different topics from politics and civil rights to health, sports and celebrity news. The website design supports the publication and consumption of original news articles and at the same time facilitates user-involvement via reader comments. Increasingly, in a period of disruptive change for the traditional media, newspapers see their future as lying in such conversations with and between readers, and new technologies to support these conversations will become essential. In this scenario there are a number of potential users: news readers and the originating journalist want to gain a structured overview of the mass of comments, in terms of the sub-topics they address and their connection with the original article and in terms of the opinions (polarity and strength) the commenters hold about these topics; news readers who join a forum discussion need to be empowered so that they can respond to the originating article (or parts of it) and/or to a sub-set of earlier comments that may be relevant to their own personal view on the matter; editors or media analysts may need a more widely scoping analysis. At present none of these users can effectively exploit the mass of comment data – frequently hundreds of comments per article – as there are no tools to support them in doing so. What they need is new tools to help them make sense of this *data deluge*.

In this section we explore in detail the categories of users who form the subject of the SENSEI Social Media Use Case (SMUC), their information needs, and the functionalities that SENSEI could provide to meet these needs. Elaborating the Social Media Use Case is challenging for a number of reasons:

- ***Novelty and Vagueness of the Task:*** No one currently carries out the tasks SENSEI will perform, so there is no question of simply automating an existing set of activities. Moreover, while it is clear that users in all the categories we have identified could benefit from improved access to and summarization of the mass of comment associated with news articles, these users have no straightforward shopping list of requirements. Indeed, most consumers or producers of reader comment have never contemplated how language technologies might facilitate their interaction with such content. Thus, asking them what they "require'', as one might proceed with requirements gathering, in more conventional software engineering application areas, is not sufficient. The response of one newspaper to a request for them to express their requirements was "We don't know. Show us what you could do for us and we'll tell you if we like it".

- ***Breadth of User Information Needs:*** We have identified seven categories of users. These user types have overlapping information needs, some specific to a particular user type, some shared with other user types. But each type also has multiple distinct information needs or interests when they approach reader comment data. The originating journalist, for example, may be interested in: the amount of response his article has generated; the polarity and degree of emotion it has provoked; any disagreements with or corrections regarding facts he has reported; suggestions for new lines for investigation in future reporting activity; etc. Supporting each of these information needs requires different technical abilities. While some of these interests may be shared with other user types,
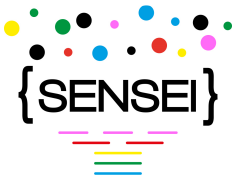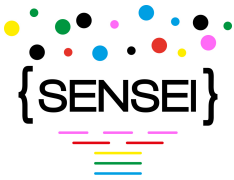
some are likely specific to the journalist alone. Altogether there is a huge breadth of potential user information interests and SENSEI has to select and prioritize which to satisfy.

- ***Technical Challenges/Limitations:*** What we might ideally like, in terms of analytics and summarization technology for the reader comment scenario, and what is possible within the timeframe of SENSEI, may well be two different things. SENSEI will push the state-of-the-art but we must be realistic about how much beyond existing capabilities we can plan to achieve. Therefore, enumerating existing language technology functionalities that may be relevant to the reader comment scenario and thinking about how to apply, combine and incrementally extend these functionalities to address user information needs is a sensible strategy. Choosing the correct balance between innovation and adaptation is not straightforward.

To address these challenges, given also the limitations imposed by the deliverable deadline, staffing and access to users, our method for determining "user requirements", or more generally "what to design and build" for the prototype SENSEI Social Media Use Case, has been as follows:

1. We familiarized ourselves with the data, collecting samples of reader comment conversations and analysing the type of language used, the relations between comments and between comments and the initial news article, and the techniques used for presentation and linking of comments on various newspaper websites.

2. We have identified **seven** types of reader comment users and have tried to articulate the sorts of information needs we believe them to have, informed by discussions with journalists and by imagining what would be useful based on an understanding of their overall role.

3. Given the novelty reader comment summarization/analysis scenario, we investigated three related scenarios – email summarization, meeting summarization and "Town Hall" meeting summarization – that we believed might give insight into the sorts of use cases that might also be appropriate for reader comment users.

4. We have identified a set of established language technology component functionalities that we believe could be useful for our user groups and act as building blocks for constructing our use cases.

5. Finally, putting together our user types, their information needs and the candidate language technologies, we have specified a number of ``use cases'' that we believe are the best starting point for the project. We adopt the definition use case given in Wikipedia: "use case is a list of steps, typically defining interactions between a role (known in Unified Modelling Language (UML) as an "actor") and a system, to achieve a goal".

Given this approach, the rest of this section is structured as follows: section 2 discusses characteristics of reader comment data; section 3 describes the user types we have identified and their information needs in relation to reader comment data; section 4 discusses related conversational summarization scenarios with a view to seeing whether aspects of these scenarios might give insights into our scenario; section 5 reviews the component functionalities on offer from language technology that may be relevant to the reader comment scenario; section 6 proposes a number of use cases that we believe are important for the users. They are technically challenging but achievable and will serve as a good starting point for SENSEI.

2.2 Characteristics of Social Media Data in the News Domain

In the Sensei Social Media Scenario we are focussing primarily on news articles that are published on-line and their associated reader comments. Our preliminary analysis has focussed on the on-line news websites of The Guardian and The Independent, the papers with whom we have links. Based on this preliminary analysis, we provide an introduction to the characteristics of the relevant news/comment data.

The typical cycle of production is as follows. A news article, written by a news professional, is published on-line and readers are invited to respond to the article, and to "join the discussion" via the comments facility, i.e. an interactive web form where registered users may add short (length limited) textual responses or *comments*. The resulting comment list or conversation is displayed below the original article. Comments are grouped into *threads*, with the comments of a single thread being presented together, in time order. When posting a comment, a user may choose to begin a new thread, or add to an existing thread, by replying to one of the comments within it. Threads often read as a series of related comments, with later comments referring back to comments that precede them in the thread. By contrast, the initial comment of a new thread may attempt to introduce a new topic, opinion or idea. This is not always the case, however; distinct threads may have similar content.

Many comments, especially thread-initial comments, may address specific content from the news article, and may express any of a range of opinions about the content, e.g., supporting or disagreeing with, adding to, clarifying, or factually contradicting. The article can thus be seen as an influence on the structure of the comments. However, as the conversation within a thread proceeds, the topic may drift away from its starting point, and new topics may be introduced. Sometimes, a comment may have no obvious direct relation to content within the article, and may appear quite unrelated to previous comments. Despite this, such apparently tangential comments may provoke many further responses and ultimately come to dominate a thread. Thus, a key characteristic of these conversations is that they are loosely focussed, with no pre-specified goal or agenda, and no active "chair" to guide the discussion.

Eventually, the editor will close the comments facility on a given news article; typically when the story becomes old news and comments have not been added for a while. The resulting data is an article and a set of comments, grouped into threads, with reply-to links between them. In addition, the may be user ratings, from users either clicking to *recommend* a comment (The Guardian), or to vote a comment up or down (The Independent).

2.3 User Categories and Their Requirements/Information Needs

We have identified seven categories of potential users of SENSEI technologies:

2.3.1 Editorial executives

Editorial executives are senior members of the news production operation. They are responsible for taking strategic decisions about the balance of content in their news publications and the deployment of scarce reporting resources. Although influenced by the paper's general editorial principles and legal guidelines, their work is also guided by economic factors, for example, they must ensure that the paper maintains and, ideally, expands its readership.
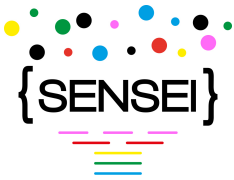
The community comments on an article, topic or range of topics, are important for news editors for several reasons. (1) Patterns in past comments may reveal community interest and preferences for different topics and reporters. Such evidence would assist editors making decisions about future content and reporting policy. (2) Comments contribute to how specific stories are received and overall to the profile of the on-line publication, which in turn affects the retention of existing readers and the attraction of new ones. Editors therefore have an interest in maximizing the quality of on-line debate.

Examples of Information Interests

- Editors may like to know which topics provoke the most comment (positive or negative; authoritative comments, etc.). E.g.:
    o Do readers want more stories from the USA, and fewer from Europe?
    o Do readers want more coverage of golf, and less of football?
- Editors may like to know which of their contributors (reporters or commentators) prompts the greatest community response (positive or negative).
- Editors may be interested in assessing the *quality* of comments and discussion for a particular story or topic over-time.

2.3.2 Newspaper advertising and marketing staff

Technology has severely disrupted the traditional newspaper business model and there is a business imperative to increase advertising revenue. In a news organization the advertising and marketing staff are interested in whether and how they can draw conclusions from on-line comments about commercial strategy, i.e. regarding how and to whom paper is marketed or how they might gather advertising revenue.

Examples of Information Interests

- Can information be extracted from reader profiles to sell to potential advertisers so that ads could be targeted at specific users when they log in? For example, on-line ads could be targeted at readers who are interesting in horse racing or cookery (sport is a good topic to address as it is easy to categorize).
- Can the general attitudes of Guardian/Independent readers be identified and be used to give potential advertisers richer information about their target audience?

2.3.3 Reporters and sub-editors

The primary task of professional news reporters is to gather information for a news story and to produce raw copy that goes forward to sub-editors. Sub-editors take the raw copy and from it produce polished output that includes, e.g. the placing of photos, the trimming of article length to fit space constraints, etc.

The extent to which reporters (and sub-editors) currently engage or would like to engage more with the content provided by reader comments is unclear and requires further study. Explicit comment or responses by a reporter in the reader comments attached to his/her article appear to be uncommon, as are acknowledgements or reference to comment within news articles. Practical constraints may have something to do with this -- our understanding of journalist's current working practice and the facilities of on-line comments, suggests they do not have much opportunity or time
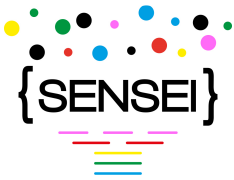
to digest the comments in any depth, nor on a regular basis. However, some studies have suggested that there are more cultural factors at play: journalists are not keen to take advantage of the opportunities presented by social media and are reluctant to surrender their role as the primary producers of news. For example, Nielsson (2013) found that "journalistic norms and conceptions of expertise" were preventing journalists from engaging with their readers. The anonymity of reader comments is an important factor here: many journalists and editors don't like the fact that posts are anonymous and they have concerns about the affects of anonymity on the quality and credibility of the comments (Nielsson, 2013; Diakopoulos and Naaman, 2011). Journalists are taught to verify their sources, and while they could test information in the comments by seeking out further references in a different context, it remains difficult for them to build a picture of a comment provider.

Nonetheless we believe journalists might have some interest in monitoring both the nature and volume of comments about specific stories on which they have been or are working. If more powerful analytical tools and summaries were available, reporters and sub-editors might exploit the comments more widely and more effectively in their work. For example, if a system were able to provide background on a comment poster, say through his previous posts reporters would have some context for assessing a comment. This might encourage journalists to consider the comments as a potential source of information, e.g., of corrections to matters of fact and ideas for future/follow-up copy. With tools to help them develop confidence in the comment posters as sources, they might also want a view of how the comments relate directly to the article content; or a view of the "emotional" impact of the story—did it polarise or unite opinion?

Examples of Information Interests

- Have comment writers spotted factual errors? -- these could be rectified in future versions of the story.

- Specific comments could suggest an angle/line of enquiry that could result in story being followed up in a particular way, either by suggesting a novel aspect or angle on the story, or by showing where reader interest appears to lie:

  o e.g., stories mentioning migrants in the UK might elicit comments about specific migrant groups (e.g. Roma, in Sheffield) of which the reporter/subeditor was unaware, prompting a follow-up story entirely on this issue;

  o e.g., a reader might say "this happened to me" , resulting in new case to report, such as financial consequences on individuals of events that happened to corporations -- these can be found only by readers volunteering accounts of their personal experiences;

- Reporter's and sub-editors might like to know if any aspect of a story elicited a strong community response:

  o A high volume of comments or highly polarised views on one aspect of the story might suggest further investigation/reporting on this aspect.

  o Stories with very few comments might suggest follow-up is not merited.

- Reporters and sub-editors would be interested in an overview view of a comment poster's (or a group of comment posters') previous postings, on different topics. In particular they might want to know if the content is:
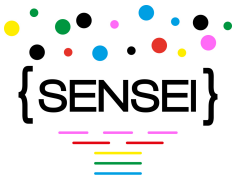
o Informed/ill-informed? Politically biased? E.g. right wing/left wing; nationalist etc. emotionally neutral or passionate? Gender neutral or specific? E.g. addressing male/female interests? Addressing youth or older interests?

This background information would help the journalist to assess the utility of a comment as a potential source.

- Moreover, the profile of posters across postings may help a reporter to assess the community as a whole: e.g. are they mainly 'sympathetic' readers of the paper vs. 'hostile visitors' to the web site (e.g. people who visit *The Guardian* site to express angry rejection of left wing views) – the paper might be differentially interested in the view of these two groups.

- An aggregation of a comment poster's (or a group of comment posters') previous posting might help a reporter to discover a novel perspective on or connection between related issues.
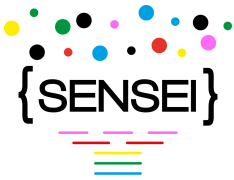
2.3.4 Comment posters

Comment posters are potentially anyone with a perspective or special concern that they wish to share in the context of a community responding to a particular article. While a common objective for a comment poster is to join a news "conversation", prior research on commenters' writing motives and our initial analysis of comments suggests that comment posters may be motivated by one or more of a variety of different factors. Diakopoulos and Naaman, 2011, deploying the *uses and gratification framework* (Miller, 2004) found evidence for "*information centric*" motives (educating others, answering/posing questions; adding information; sharing experience, clarifying, note missing information, balance discussion, correct inaccuracies or misinformation); "*personal identity*" motives (emotional factors such as expressing intense feelings, or opinion); "*entertainment*" motives (inject humour into discussion); and "*social interaction motives*" (see reaction of community, persuade others, sympathise, leave condolences, applaud goodness, debate etc.); and we believe there may be many more, e.g. developing background interest in a topic; political factors (e.g. a desire to participate in debate or "democratic" exchange; to show/encourage support for a political group or cause), etc.

While the current facilities for comment posting enable debate and interaction to some extent, there are various limitations and challenges. Two key issues include:

- The prevalence of "off-topic" comments and large volumes of comments -- if a comment poster has a specific concern and they wish to engage with others who have something to say on the issue, how do they find other relevant comments or make their own comment more visible to interested others? How can a comment poster find the best point for him to join the conversation more quickly?

- The posting of abusive comments – if this practice is widespread such comments may deter some people from posting; on the other hand, excessive moderation may deter comment posters, on the grounds that the forum is not "democratic".

Examples of Information Interests

- Comment posters might like to see and have access to a breakdown/grouping of existing comments in relation to the aspects of the original story that they address and in terms of other topics they address.
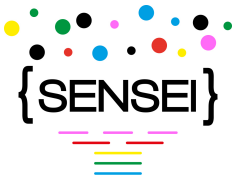
- Comment posters may wish to get a summary of the comments in a thread, both in terms of content and in terms of polarity in sentiment.

- Comment posters may like to have a profile of other posters in a specific conversation and in the wider news community, e.g., who's active in the conversation; are they active members in the community; who is talking to whom and on what topics; what are their interests and preferences, as indicated by their previous postings, etc.

- If the system can deliver a (balanced) summary of comments, posters may wish to comment on that in turn, i.e. summary becomes a story in itself; e.g. if comments summarised as "commenters were in favour of UK becoming a republic".

- Comment posters might like a tool to filter "authoritative" or "key" comments in the conversation.

- When deciding whether to join a conversation, comment posters might like to see an assessment of the overall "quality" of a thread or conversation.

2.3.5 Comment readers

Comment readers include public consumers of on-line news who have an interest in finding out about the community response to a particular article or story or topic, at some point in time and/or, possibly over-time. We believe that this group may include both (1) people interested in a news story, who are interested in seeing if the community has anything to add to the story; and (2) people who have an interest in learning more about the opinions of the comment poster community on topical issues, and possibly about individual comment authors, because they find them interesting or want to see where they may be coming from, i.e. what has a poster said about similar or other topics? Previous work suggests that more can be done to encourage more people to read on-line news comments, and/or to enable existing "comment readers" to do so more regularly and in greater depth (Diakopoulos and Naaman, 2011). Particular challenges include the sheer volume of comments; the perceived quality of comments and prejudice towards comment poster communities, the lack of indexing and navigation/search facilities and abusive comments.

Examples of Information Interests

- Comment readers may be interested in a balanced summary (of topical content and/or polarity/and volume;) of comments, given lack of time/motivation needed to read all of them.

- Comment readers might like to see a breakdown of existing comments in terms of how they address or extend aspects of the original article.

- Having found a comment that interests them, comment readers might be interested to see a profile of the comment poster based on their previous comments. This may help the reader to assess the comment poster's current contribution.

- Comment posters may like to have a profile of all posters in a specific conversation, e.g., who's active in the conversation; are they regular comment posters; what are their interests and preferences, as indicated by their previous postings?

- When deciding whether to read a thread or conversation, comment posters might like to see an assessment of the overall "quality" of a thread or conversation.

- Comment readers might like a tool to filter "authoritative" or "key" comments in the conversation.

- Comment readers might like to have a summary or overview of the comments posted in relation to multiple stories in a topic.

2.3.6 Reader comment monitors/sub-editors

As reader comments to on-line news have grown in number and popularity, newspapers are beginning to employ staff whose dedicated role is to monitor, promote and engage with on-line comment. For example, in the case of *The Guardian,* for all articles in the paper on which comment is invited, staff identify reader comments that they believe contribute significantly to the debate and give them a "Guardian Picks" seal of approval. Additionally, *The Guardian* is promoting, via their on-line "Comment is Free" page, opinion pieces by Guardian staff and free-lance authors, approved content from volunteer contributors and selections of reader comment on these opinion pieces. These features show how the paper is taking seriously the idea of involving readers in the production of content and shaping of debate.

We believe tools that could automatically suggest high quality reader comments could be of interest to such comment monitors/sub-editors, reducing the burden of reading huge numbers of low interest comment. In addition, although at present comment monitors/sub-editors do not attempt to characterise the entire set of comments in response to an article or set of articles, a "round-up" of social media comment, such as is becoming commonplace on television and radio programmes (e.g., Twitter #olympics) is something that could be of interest, if technology were in place to support it.
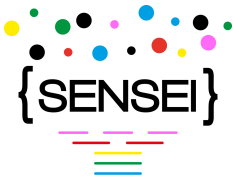
Examples of Information Interests

- Comment monitors/sub-editors may like to have a candidate set of quality comments automatically selected for them from which they can select comments to recommend.

- Comment monitors/sub-editors may like to have a summary picture of all comment for a day to help them
    o identify the most active, polarized or emotionally charged debates; and
    o prepare a round-up type summary piece on the day's comment.

2.3.7 Media analysts

Aside from those involved in the production of news and those reading or commenting on it, there may be 3rd parties who are interested in analysing reader response to news stories, i.e. media analysts. These might be organizations with a direct interest in the information or intermediary organizations (e.g. companies specialising in media analysis) hired by other organizations to gather this information. There could be a new market in newspapers selling this information to other organizations.

In the past decade academics in the social sciences (working in areas such as communication and political science and journalism studies) have carried out studies of reader comments, reporters and comment posters. While this work has yielded some important insights, the studies have typically been limited in scale due to the technological challenge of processing large volumes of social media content (Ruiz et al., 2011; Manosevitch and Walker, 2009; Strandberg and Berg, 2103). We believe that academics interested in studying the character of public exchange and

debate would be interested both in using text-analytics in their work and in studying the effects of text analytics on the conversations.

Examples of Information Interests

- Companies might want to know the profile of their company/product in social media.

- Political parties or polling organizations might want to take the temperature on particular issues.

- Social scientists or researchers might want to study public reaction to various news stories.

- We believe academics in fields such as communication and political science and journalism, would like to have various kinds of analyses of large sets of reader comments over-time, to acquire background for studies investigating the character of on-line news conversations. Specific interests might include: measures of political preferences – are comments expressing left-wing/right-wing views?; a record of polarised sentiment on particular topics; measures of diversity within reader comment communities; measures of topical diversity in the comments; how much of the comment relates directly to article content?; what proportion of the comments may be viewed as "low-quality", e.g. off-topic; nonsensical, etc.; how prevalent are abusive comments?

2.4 Conversation summarization: related scenarios

In the previous section we have described user needs based on discussions with journalists and technologists. Here we discuss related conversation summarization scenarios with a view to seeing whether aspects of these scenarios can give insights into our scenario. We discuss three related scenarios: email summarization, meeting summarization and what may be called "Town Hall Meeting" summarization.

2.4.1 Email Summarization

One scenario that could offer insights into what information users would find useful in summarizing readers' comments is summarizing e-mail threads. This scenario has several features that make it similar to the reader comment case, specifically: (1) the conversations may be multi-party; (2) the language is colloquial and includes lots of abbreviations, spelling mistakes, etc. (3) conversation may drift away arbitrarily away from the initial topic. However, there are several aspects of email and reader comment summarization that are very different: (1) in the reader comment case there is an initial news report and comments need to be interpreted in relation to it; (2) emails, when they are replies to an earlier mail, frequently copy in large chunks of one or more earlier messages which serve to structure the reply and suggest topical focus, while reader comments do this to a much lesser extent; (3) external information, such as thumbs up or down in relation to comments, or profile and previous posting behaviour of conversants is not available; (4) email summarization is typically carried out at the level of individual messages or threads of messages replying to an initial message, while reader comments typically take the form of multiple threads, each consisting of multiple comments; i.e. the scope of the summarization required is broader.

There has been some work on automatic email summarization, but little discussion of user-related issues, e.g., specific user task contexts in which email summarization would be useful or task-centreed evaluations (one notable exception is Corston-Oliver et al. (2004) who consider the task of identifying potential action items from emails for the reader to consider adding to their "to do" list). For the most part the assumption by the language processing community has been that
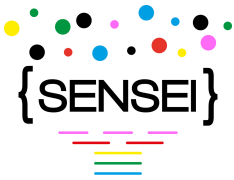
generic informative summaries of email threads would be useful things to produce. Evaluations of developed systems (e.g. Rambow et al. (2004), Carenini et al. (2007), Murray et al., 2008) tend to focus on correct identification of sentences deemed important, by human assessors, for inclusion in an extractive summary, where the human judgements have been obtained without reference to any task context. In addition to gathering reference data for evaluation, this set of human generated e-mail summaries also offers an insight into the information needs of the potential users of e-mail summaries. In the case of Carenini et al. (2007) the interesting finding is that only 12% of sentences in their data are agreed upon as essential by human summarisers. Thus, although people may have a general need to read e-mail summaries in order to speed up their information access, there is no unanimous agreement on which information from the e-mails would best address this need.

The differences between email and reader comment summarization, the broadly accepted unexamined presumption of the utility of generic informative summaries and the lack of agreement of human subjects in selecting core content for inclusion in a generic email summary, suggest that work on email summarization is unlikely to be very helpful in defining use cases for reader comment summarization/analytics (of course the techniques developed to carry out email summarization may prove useful for SENSEI).

2.4.2 Business/project meetings

The second scenario we consider is summarization of spoken multi-party interaction as presented in the work on a meeting browser within the AMI project (Renals 2004).
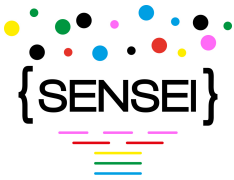
Methodologically, the study presented by Tucker et al. (2005) is interesting. For defining use cases for the development of the meeting browser the authors investigate two methods for eliciting users' needs. First, the users' current practices in dealing with meeting data are be explored. Second, the users are confronted with a new, non-existing technology, like having a meeting summary, to assess their acceptance of it and observe the way in which they deal with it. The authors argue that a combination of both approaches is essential to arrive at conclusions about users' needs. The users' current practices can be collected via working prototypes, while new technologies can be introduced using paper-based prototype simulations.

2.4.3 The Town Hall Meeting Scenario

The town hall meeting is a traditional form of social dialogue where members of a community have the opportunity to come together to discuss a particular event or proposal and to exchange ideas and views among themselves and with elected officials, political candidates, public officers and representatives of municipal organisations. (Indeed we can find evidence for such meetings taking place in Medieval and even Roman times.) Common examples today are local council meetings where officers present and gather views on local issues such as flooding, planning etc. Other examples include meetings where politicians, such as councillors, MPs and Mayors, engage with an invited audience, for example drawn from a constituency or school population, in a question and answer session, which may cover local and national issues. Public officers and members of the media may take records of such meetings. News reporters may take notes, write summaries and publish reports. More newsworthy events may be covered by radio and television crews.

2.4.4 Similarities Between Town Hall Meetings and Reader Comment Conversations

We have identified a number of important similarities between the town hall meeting and news reader comment conversations. Both typically involve members of the public raising issues and

responding to each other in a loosely focussed, social, dialogue. The town hall meeting often has a central theme, i.e., it is about a particular topic or issue, which may be presented verbally by a speaker at the outset of the meeting and this theme may play a similar role to that of the original "news article" in the reader comments scenario. The town hall meeting and the reader comments scenario are not typically motivated or guided by a set of pre-specified goals or actions, e.g. a vote; nor must they adhere to a strict agenda requiring actions at different points in the meeting (more typical of, say a business or project meeting). And, while the direct questioning of public figures is not a typical characteristic of reader comment conversations, we do see comment posters enter into question and answer style responses with other posters. A major difference is that comment posters may choose to conceal their identity from fellow posters and ultimately all communication is via the written word; by contrast, in a town hall meeting members of the audience will be visible to others present at the meeting, and although they can still choose to reveal very little about themselves, the fact they are "face-to face" with their audience is likely to have an influence on the nature of their spoken interactions. Another difference is the time frame in which views may be exchanged: a town hall meeting is more time constrained than an online conversation and people attending a town hall meeting may feel they have more limited chances to get their views across, whereas in social media they may keep on posting indefinitely.

Nonetheless, with such striking similarities in the character of the conversations, we can view the task of a reporter taking notes and reporting on the town hall meeting as analogous to the SENSEI task of producing summaries of conversations in the news domain. This analogy is important, since it provides us with a real world example of conversation summarisation, which we can study in various ways in order to obtain insights into what social media summaries could look like2. For example, we can examine reports of public meetings, such that shown in Figure 1, Appendix 1, for evidence of content selection and examples of how content is presented. Alternatively we can ask journalists about their current practice: how they make notes and what they are looking for when they report on such meetings. We can also seek out records or minutes of meeting content and then explore how the dialogue content relates to the reporter's notes and the summary news article. And finally, we might construct simulated town hall meetings, perhaps based on dialogues constructed from social media, and then invite people playing the role of reporters to make summaries of the meetings. Once we have gained an understanding of what the summary should be like, we can proceed to develop use cases based upon the town hall style summary (see Section 6 below).

We now present results from initial research into town hall summaries, based on discussions with journalists at the University of Sheffield, and a preliminary analysis of news reports of summary meetings. In the next sub-section we describe the kinds of content a reporter would be looking to include in his notes and final report. Appendix 1 shows a news report of a town hall meeting, and examples of how the content types we list below appear in the report.

2.4.5 A "Reporter Protocol" for Summarising Public Meetings
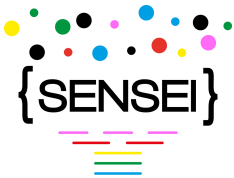
An outline of the main content a reporter might seek to include in a summary of a town hall meeting is as follows. He/she would not necessarily include all of this content in their notes or final report:

---

2 The analogy also provides us with a useful real world context for grounding and focussing evaluation tasks, and as such is useful for future deliverables and work in SENSEI.
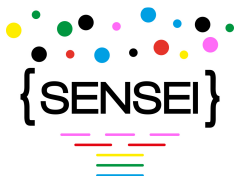
the list is simply indicative of the types of content a reporter would consider. Moreover the content is often determined by what actually happens at the meeting and the information available.

- **Meeting Context**:
  - What the meeting was about and/or why it was called
  - A "headline"- i.e. an overarching summary or "angle" on the meeting (a short 1 or 2 line summary, summing up the reporter's view on the meeting)
  - Background to the meeting
  - Where the meeting was held
  - Time & date of meeting; how long did it last

- **Who attended**:
  - How many attended
  - Who were the official speakers (e.g., Mr. Mayor, Councillor Smith, Officer Jones, MPs)
  - Who was in the audience – the aim here is (i) to provide a profile of the audience, for example in terms of: local people/outsiders/mostly male/mostly female/old/young/political affiliation, etc. (ii) to identify notable individuals

- **Meeting Content**: An account of the discussion in terms of topics or issues covered and the opinions voiced on different issues. This might be done through the following.
  - General observations
    - The main issue under discussion was/ most of the discussion focused on […]
    - The topics covered at the meeting were/ranged from […]
    - Other key issues were […], and who they were put to.
    - There was consensus on issues […]
    - Opinions were divided on issues […]
    - There was strong support, by whom, for issues […]
    - There was strong criticism, by whom, of issues […]
    - Opinions ranged from […]
    - The most widely held opinion/view was […]
    - A minority of people were of the view/opinion […]
  - Selected quotes
    - The most salient or most compelling comments
    - Comments which help to define the boundaries of the discussion (i.e. which help to delimit the space of comments and opinion)
    - Comments which exemplify the discussion of key issues

- **Meeting Outcomes**:
  - The most striking outcome
  - Other notable outcomes

2.4.6 Future Work –Towards a Model Town Hall Meeting Summary

In future work we could obtain a more detailed model of how reporters summarise town hall discussions. For example, much more could be learned from a systematic study of such news reports: e.g. the nature of rhetorical relations within the end summary; the order of information presentation etc.

2.5 SENSEI functionalities for the social media scenario

As noted in the Introduction of Section 2, part of our methodology in arriving at use cases is reviewing the set of technologies – both language processing technologies and non-language processing technologies – that can be brought to bear in the reader comment scenario. While SENSEI intends to push beyond the current state-of-the-art, we cannot expect to solve all problems in computational language processing and need to build realistically on what is available. Thus, current functionalities both support and constrain efforts in SENSEI. We need to know what these are to specify use cases that incorporate them.
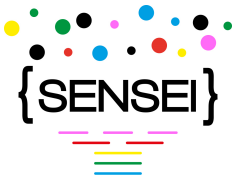
2.5.1 Statistical Profiling

By "statistical profiling" we mean the process of gathering basic quantitative data about one or more conversations or one or more comment posters. Techniques to do this are quite straightforward and do not involve language processing. For a single conversation data could be gathered about the following.

- total number of comments
- total number of commenters
- number of threads
- maximum, minimum and mean number of comments per thread
- maximum, minimum and mean length of comments
- maximum, minimum and mean number of responses per comment
- total number of comment "approves", maximum number of comment approves for a single comment

Similar statistics could be aggregated across multiple conversations (e.g. on a topic or over a time period) or contrastive numbers could be generated for any pair of conversations or pairs of sets of conversations (e.g. on different topics or in different time periods). Conversations could be ranked according to any of the above statistics (such as total number of comments).

Statistics can also be collected for individual posters or groups of posters. For example, for a single poster data from their profile can be extracted to determine

- total number of comments they have posted
- total number articles they have commented on/threads they have participated in,

- total number of "approves" their comments have received,

- total number of posters they have replied to/number of posters who have replied to them,

- maximum, minimum and mean number of exchanges with a specific set of other posters.

The above data could be presented in time slices as well, to get a picture of activity over time.
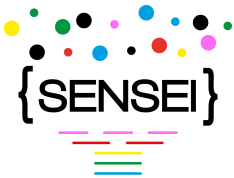
2.5.2 Topical Analysis and Linking

We consider aspects of functionality relating to the topical content of text and speech. There is no fixed or agreed definition of what constitutes a *topic*, but in certain contexts we intuitively recognise that a particular sequence of material constitutes a coherent topic, and that moving to the material that follows constitutes a change of topic. For example, in a news broadcast, we might see the different news stories reported as constituting a series of separate topics. In a transcript of a formal meeting, the different topics might correspond to the agenda items. However, issues of *granularity* arise, and what is viewed as a single coherent topic at one level might, on closer inspection, be seen to involve a series of subtopics that are addressed in turn. The discussion of a single meeting agenda item, for example, might pass through several phases, as various relevant issues are considered. Likewise, whilst a news article might be seen to have a single overall topic, one succinctly expressed by its headline, the paragraphs of the article may be seen to address a series of topics that fit within the encompassing topic of the article. In what follows, we list relevant topic analysis functionalities. Despite the differing tasks, most approaches have a common underlying idea that a sharing of topic between texts (or text fragments) is reflected in the words – or, more abstractly, concepts – that appear in them, and in the entities to which they refer.

Topic Segmentation: This functionality divides a text into topically coherent segments. The text in question might be a written text, such as a news article, or may be a conversational exchange, such as a transcript of a spoken dialogue. Most segmentation approaches are *linear* in characterising a text as a single flat sequence of topic segments, although a *hierarchical* analysis is also possible, reflecting the issue of granularity mentioned above. For asynchronous conversational media, such as news comments and twitter exchanges, the conversation may interleave discussion of more than one topic, which complicates the issue of what constitutes a topical segment. One of the high-performing approaches to topic segmentation is based on recognising lexical chains, which capture patterns of term repetition across the text.

Topic Labelling: This is the task of providing a short description of topic segments, to facilitate their interpretation. Such descriptions may consist of a collection of key words or phrases, which might extracted from the text, or arrived at in a different way, such as by selecting the top-ranked terms of a topic model (such as those produced by LDA topic modelling).

Topic Clustering and Linking: A key basic functionality is being able to recognise *topical similarity*, i.e. that two text fragments have similar content, from which we might infer a shared topic. This ability has been used as a basis for clustering documents within a collection as belonging to the same broad topic. It might also be used within a document as a basis for determining that non-adjacent topic segments are topically similar. A particular case of this of interest for SENSEI, that arises for news+comment data, would be linking a comment, or group of comments (which form a topic segment), back to a topic segment within the news article, as a basis for saying that the comments are *about* that part of the article. The identification of such *aboutness* links between comments and the article would contribute to the interpretation of comments, but a more elaborate subclassification of such links can be envisaged and may also be

useful, e.g. indicating that a comment *agrees with* an article segment (or another comment), or *disagrees with* it, or perhaps that it *clarifies*, *extends* or *contradicts* it.

2.5.3 Sentiment analysis

Sentiment analysis covers several tasks:

- Opinion extraction: identifying a span of text that contains an opinion.

- Opinion classification or orientation: determining whether the opinion is positive, neutral, or negative, and perhaps indicating intensity on a numeric scale. This can be carried out using machine learning techniques (as a text classification problem) or with rule-based systems using gazetteers (such as data extracted from SentiWordNet). Detecting sarcasm is difficult, however, and poses an interesting current research problem.

- Target identification: identifying what the opinion is about. This may involve associating parts of complex opinion statements with specific targets (e.g., "The picture quality [of a camera] is good but I'm not happy about the battery life."). For comments about news, the results of named entity recognition are often a good basis for identifying targets.

- Source identification: determining who holds the opinion. For short texts this is usually the author of the text, but in some cases, this task requires resolving co-references and indirect speech.

Target identification can be tied in with the results of topic extraction, although user-generated content, such as comments on news articles, is subject to considerable thread drift so it will be necessary to identify deviations from the main topic. This issue can also be linked to authority analysis (in the next section)—commenters who frequently wander off topic, especially if they try to bring many discussions around to their own "pet topics", may be regarded as less authoritative.
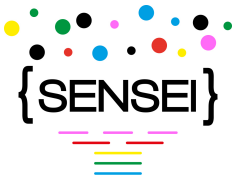
Opinion classification with numeric scores can also be used to produce interesting statistical work, such as measuring the "interestingness" of opinions in terms of how far out of the "middle of the road" their scores are. On the other hand, the more moderate opinions might be associated with more authoritative comments and commenters (see below).

2.5.4 Authority Analysis

What *authority* means exactly is debatable and differs slightly between disciplines. In our scenario, authority assessment of both the comment posts and the comment posters (which may converge or differ from post to post) is needed in order to indicate reliability or trustworthiness of the information in the posts. The comment posters' authority rating can be used in selecting the comments or simply for providing information to the end user about the authority rating of the sources that the extracted information comes from. How can we measure the authority of comment posters and their posts?

2.5.5 Extrinsic Assessment of Authority

Several measures for assessing authority can be computed from features found in online newspaper comments that are extrinsic to the conversation itself. Some, like *number of recommendations* in *The Guardian* or the option to *report* offensive content, rely on community assessment. Others, like *staff replies* and *Guardian picks*, are provided by the newspaper itself in an attempt to foster useful online discussions. Furthermore, statistics like a user's activity on the forum or number of responses to or from other users could be combined with their community or

staff assessments and mapped to their authority. They can also be used as training classes for machine learning (text classification).

However, the reliability of all these extrinsic assessments is questionable. The very same values of these attributes can often be linked to authority in both positive and negative sense. For example, a user who is very active on the forum and receives a lot of thumbs up or recommendations, which could suggest "positively authoritative", can also have a high number of deletions for offensive contents. This has been observed in highly polarized political discussions for example, where often the same users repeatedly take part in discussions, join one of two or more opposing opinion groups and give positive assessments within their own group and negative to the opponents. Additionally, there is a risk that these ratings systems diminish quality by rewarding "groupthink" and encouraging users to make comments that they know will gain many other users' approval.

Newspaper staffs' choice of online content based on their own subjective authority assessment has been questioned too. The main criticism is that it can be seen as a form of censorship to highlight contributions that an institution holds to be good, important or true and therefore affect free opinion formation among online readers[3]. The newspapers' invitation to comment and join the discussion would then become an invitation to join the discussion that suits them.

2.5.6 Intrinsic Assessment of Authority

A further way to assess the authority of the readers' contributions and of the readers themselves based on their contributions is to infer an authority rating from the text of the posts. This is a challenging task. There are several possible feature sets that could indicate the posters' authority in wider sense, but all need further investigation in relation to establishing authority. Features of the text, like correct spelling or use of grammar or certain lexical choice could all potentially be positively mapped to authority in that they estimate the educational level and sophistication of the writer.
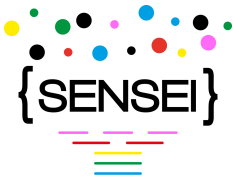
Previous research suggests that certain types of conversational actions can be positively linked to authority. In a study of task-based spoken dialogues Mayfield et al. (2011) establish that seeking information (by asking a question for example) is related to low authority, while sharing knowledge or giving instructions links to high authority. In order to cross-verify this, the authors report a study of the relation of their authority assertions to certain emotional states, like aggression, in which higher authority positively correlates with higher aggression. This indicates that, possibly, assigning the posts an emotivity index would be helpful in authority assignment. However, this needs further work to see whether and in which direction the correlations may exist.

Furthermore, the results of sentiment analysis inform authority assessment, although this also needs further investigation and validation. For example, opinionatedness could be negatively linked to authority: opinion scoring (see section 2.5.3) could be used to mark extreme opinions and their holders as less authoritative.

2.5.7 Self and Other Initiated Assessments of Authority

Finally, one can draw on peoples' self-assessments and other-initiated assessments in other to establish authority. Bender et al. (2011) show that forum posts often come with claims of authority

---

[3] http://www.bookbusinessmag.com/article/open-annotation-comments-hypothesis/1#
www.hypothes.is

made by the poster, e.g., "I have 20 years of experience in X and I say….". They propose a classification scheme in which authority claims are placed into 6 categories. Since these are lexically well defined, it may be fruitful to use lexical features for classification of readers' comments into one of these categories.

Other-initiated assessments are also a possible source of authority claims. In this sense, each comment's authority index would be established based on whether and how it is talked about by other conversation participants. This is similar to the method of Conversation Analysis (e.g. Drew 2005), which is typically used for spoken interaction.

Other-initiated assessment could be based both on deeply analysing the text of co-commenters. However, it could also be implemented as an interactively generated authority score between a comment poster and those in the community who link to him/her. Given some other indication of authority, or just starting from a baseline where all users are given equal ratings, an algorithm similar to Google's page rank could be used to increase the authority scores of users according to the authority scores of those who respond to them and those whom they respond to.

2.5.8 Summarization

Automatic summarization has been one of the central subtopics of applied natural language processing since the mid-1950s. Work in this area may be positioned along three dimensions: (1) the type of content being summarized (2) the type of summary being produced and (3) the technique(s) being employed to generate the summary. A good recent overview of automatic summarization may be found in Nenkova and McKeown [2011].

Type of Content/Input

The earliest research on automatic summarization addressed the problem of summarizing single written documents, e.g. generating an abstract for a scientific paper. However, the rapid growth in availability of digital news and the "information overload" problem has led to increasing interest in multi-document summarization. Here the task is to generate a single summary that merges the information from separate, presumably topically related documents and selects the most central content, eliminating any redundancy, e.g. summarizing multiple accounts of the same news event. Well-known systems that address this task are MEAD [Radev et al., 2004] and NewsBlaster [McKeown et al., 2002]. Recently the problem of summarizing multiple on-line consumer reviews (e.g. of restaurants) has begun to receive attention [Carenini et al., 2012; Di Fabbrizio et al., 2011]. Summarizing such reviews raises new challenges, specifically how to summarize opinions and how to deal with the noisiness of on-line language. Unlike news, where repetition across multiple articles needs to be disregarded, in review texts repeated expressions of similar opinion are essential in building up an overall picture of customer sentiment regarding a product or service. Furthermore the colloquial, frequently ungrammatical and misspelled or abbreviated nature of the language used in on-lines reviews makes them particularly challenging. Different sorts of challenges are raised in summarizing spoken language and in summarizing dialogues or conversations, challenges that are combined in the case of summarizing spoken dialogues, such as speech recognition errors, handling disfluencies, sentence boundary detection and exploiting prosody [Zechner, 2002a]. Dialogues, whether spoken or written, raise other issues such as cross speaker information linking (e.g. question/answer pair detection) [Zechner, 2002b]. Researchers have explored summarization of meeting transcriptions [Murray et al., 2005; Murray and Carenini, 2008], telephone call or contact centre conversations [Byrd et al., 2008; Higashinaka et al., 2010;
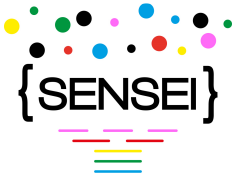
Tamura et al., 2011], email threads [Murray and Carenini, 2008], on-line chat [Zhou and Hovy, 2006; Uthus and Aha, 2011] and reader comment in newspapers [Ma et al., 2012].
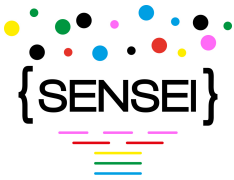
Type of Summary/Output

Another dimension of variation in summarization research is the type of summary produced. Distinctions here include:

- *Indicative vs informative summaries*: Is the purpose of the summary to enable the reader to decide whether the original content is relevant to his/her task, and hence worth reading in detail, or to serve as a surrogate for reading the original?

- *Guided or query focused vs generic*. Is the summary meant to address particular concerns of the reader, e.g., including key aspects of certain event types such as location of a natural disaster? Or is it to capture the core information for a generalist audience?

- *Extractive vs abstractive*: Is the summary comprised of sentences selected from the original documents/speech? Or does it consist of new language generated by the summarizer from some intermediate representation computed from the inputs?

- *Graphical vs textual vs multimedia:* Does the summary consist solely of text in the form of a coherent text? Does it consist of keywords/snippets? Is it in graphical form? A mixture of the graphics and text? Does it contain pointers into the text or speech stream it is summarizing?

Summarization Techniques

A final dimension of variation in work on summarization is the techniques employed. These depend on the type of summary being generated, particularly whether it is to be extractive or abstractive. Approaches to extractive summarization rely on ranking input sentences in terms of their importance for inclusion in the summary. This is done by extracting features considered summary-relevant. These can range from basic word-related features, such as lexical overlap with title words, presence of cue words, similarity of sentence to document or document set centroid [Radev et al., 2004], to more sophisticated features involving, e.g. coreference chains and discourse relations [Nenkova and McKeown, 2011]. From the extracted features an overall sentence score is computed typically as a weighted linear combination of these features, where the weights are either heuristically determined or learned from data.

By contrast, abstractive approaches do not simply select and order sentences from the input documents. Rather, they first perform some analysis of the input and then, based on this analysis, generate new sentences distinct from the input, though possibly edited versions of the originals. While in theory abstractive approaches could involve full text understanding, followed by generation, in practice, since full text understanding is still well beyond reach, abstractive approaches tend to follow one of two simpler approaches. First there are those that, because they address a domain-specific summarization task, rely on information-extraction-like techniques to map surface texts into pre-defined domain-specific template structures that capture core information elements in the domain (e.g. McKeown and Radev (1995)). The second group of approaches directly address the weaknesses of extractive approaches, specifically: (1) extractive approaches include or exclude full sentences only, while in many cases only part of the information in the sentence may be summary-worthy; (2) in some cases information elements are repeated across multiple selected sentences; (3) extractive summaries may be incoherent due to inclusion of anaphors without antecedents or infelicitous ordering of content. These problems can be

addressed by systems that work in conjunction with an underlying extractive approach. They have been addressed through work on sentence compression, information fusion, sentence revision and information ordering [Nenkova and Mckeown, 2011].

Hybrid extractive/abstractive approaches, such as that proposed in Fabbrizio et al. (2013) for summarizing restaurant reviews, which combine information about the distribution across reviews of opinions on aspects of the restaurant under view (price, quality, service, etc.) with selected quotes, are especially relevant to SENSEI.

2.6 SENSEI Use Cases for the social media scenario

Putting together

- our analysis of user groups and their information needs from Section 3,

- insights from related summarization scenarios discussed in Section 4 and

- the capabilities and limitations of available analytic and language processing technologies described in Section 5,

we are now in a position to propose a set of use cases that we believe address the information needs of SENSEI users and that, while technically challenging to realize, are achievable.

In the following we present six candidate use cases. The first two are digests of the entire set comments related to a single news article. The second two start with a single reader comment and look at building context to help users interpret that post. The final two take multiple conversations (set comments related to a single news article) as inputs and provide meta-level digests or analyses.

In terms of the summarization types presented in the original SENSEI proposal, Use Cases 1 and 2 are forms of "conversation-oriented summaries" for single conversations; Use Case 3 is a form of "blogger-oriented summary"; Use Case 4 is a type of "ad hoc summary"; and Use Cases 5 and 6 are forms of "conversation-oriented summaries" of multiple conversations.
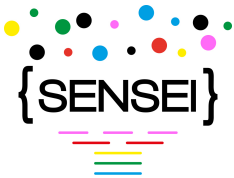
2.6.1 Use Case 1: Town Hall Summary

Actor: Reader, reporter, comment poster, comment page editor, social scientist.

Goal: To derive an overview or characterisation of the set of reader comments attached to a particular article, such as would be obtained from a news report of a Town Hall meeting, as described above in Section 4. This overview should include: a brief summary of the article the conversation was attached to, a brief over-arching summary of the conversation itself, a profile of the social dynamics in the conversation (e.g. the key contributors to the conversation; and the pattern of posting e.g.: were the comments dominated by a small number of frequent posters or were many different people commenting with single posts, etc.?), the scope of the conversation in terms of the major topics addressed in the comments and the volume of comments associated with specific topics; the polarity and intensity of feeling associated with particular topics.

Steps:

1. The reader requests a town hall summary for a news article and the set of associated reader comments on the article.

2. The system analyses comments and article and:

- provides a brief (e.g. 1 line) summary of the article content and comment conversation;

- determines the number of individual comment posters in the conversation and assigns an indication of their comment history (e.g., "very frequent poster", "occasional poster", "new poster");

- determines the structure of the threads (e.g. total number of threads; the average number of comments per thread; the number of 'unique' comment posters per thread);

- ranks the contributing comment posters (e.g. based on information from the total number of posts, initiated threads, number of direct replies to their comment in a thread, total length of posts);

- determines major topics addressed in the article and comments;

- determines polarity and intensity of comments;

- aggregates numbers of comments per topic and numbers of comments per sentiment category;

- selects illustrative quotes from "authoritative" comments to exemplify views and sentiment with respect to topics included in the summary.

3. The system generates/presents a summary report that includes all of the above.

We note that in an ad-hoc scenario, the user could specify all or any sub-set of the above summary content.

An example of a Town Hall-like social media summary has been included in Appendix 1.

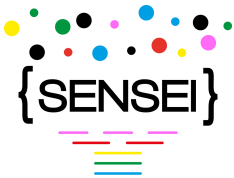### 2.6.2 Use Case 2: Identifying Comments That Extend the Coverage of the News Story

Actor: Reporter; Comment Reader

Goal: To gather from the reader comments any additional information that refers directly to the article content, including: factual errors; additional facts/elaboration; context; personal experiences and recommendations for follow-up; with a view to either (1) (reporter) updating the article or providing content for a follow-up story or (2) (comment reader) extending personal knowledge or understanding of the story. Note that this Use Case excludes assertions of opinion or off-topic comment.

This use case may be specified in generic mode or ad-hoc mode:

Steps: (generic mode)

1. The journalist/reader identifies the article to the system.

2. The system analyses the comments in relation to the article and gives a breakdown of the following information.

- Reports of factual errors.

- Comments which elaborate on any content in the article e.g. by introducing new, or related facts or evidence.

- Comments which propose similar examples to article content .

- Any accounts of personal experience of issues raised in the article.

- Recommendations for further, related lines of enquiry.

In ad-hoc mode the journalist/reader identifies the article to the system, and selects one or more aspects of the article in which he is interested. He then selects one or more options from the above list to include in the system report (e.g. "factual errors") and then the system analyses the comments and provides a user-tailored report.

We note that when examining and assessing the utility or authority of content provided by the comments the user might choose to find out more about the comment provider, especially in respect to his past comments on other articles (see Use Case 3) or he may wish to examine any similar comments made by others in different contexts (see Use Case 4).

2.6.3 Use Case 3: Backgrounding: Looking in Greater Depth at a Comment Poster

Actor: Comment Reader/Comment Poster/Reporter

Goal: To find out more about a comment poster whose comment on an article is of particular interest to this reader, and/or with a view to gaining a measure of confidence in the value of a comment (Is this comment likely to reflect certain interests; have any substance or truth?).
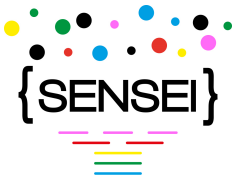
Steps:

1. The reader identifies the poster-of-interest to the system by selecting a particular one of their comments

2. The system returns a report to the reader that includes the following:

   - whether the poster of this comment is a prolific commenter;

   - the range of subject areas across which this poster comments;

   - whether this poster's comments typically garner many responses, and whether the responses tend to be positive or negative;

   - whether this poster exchanges comments with many other posters, or whether s/he tends to interact with a limited set of other posters, and if so who;

   - a characterization of the poster and their interests in terms of the language they use, e.g. a word cloud or key phrases.

3. The system also provides an interface that allows the reader to `drill down' to see other comments by this poster, within the conversation and in other conversations, either in context or not, and potentially filtered by subject area, interlocutor, recency, etc.

2.6.4 Use Case 4: Finding Similar, Related or Redundant Postings

Actor: Comment Poster/ Reporter/Comment Reader

Goal: given a comment that either (1) a reporter, reader, or poster is interested in or (2) a poster is currently authoring, the goal here is to find out whether there are other comments, already posted, that make the same point or are closely related in content. The user may choose to search for similar comments either within a particular conversation or from other conversations. He carries out his search with a view to either (1) (reporter/reader/poster) assembling context in order to gain a measure of confidence in the value of a comment (is this comment saying something other people have said before? Is this comment likely to have any substance or truth? Am I right in saying this?

or (2) (comment poster), identifying redundancy (do I really want to make this point, if others have already made it many times before?)

Steps:

1. The user provides their comment of interest (either an existing comment or a draft of an intended comment) and clicks to check for closely related postings, either in the context of the same article, or in reader comments in related topic areas (e.g. "other environment stories"; etc.).

2. The system returns a list of comments, selected from those already posted on the article, or other related articles, ranked by their degree of similarity to the comment of interest.

2.6.5 Use Case 5: Identifying Trends in Reader Comments

Actor: Editor/Media Analyst/Social Scientist

Goal: To determine which news stories/topic(s) in a specified time period (e.g. week, quarter; year etc.) have elicited a significant response from the comment posting community. Key trends will include: stories/topics with very high/low volumes of reader comment; stories/topics with the most emotive reader content. Trend seeking is carried out with a view to steering the paper's content policy (editor), or with a view to gauging public opinion (media analyst), or with a view to examining patterns in public conversation and debate in the on-line news domain (social scientist).

Steps:

1. The editor/media analyst/social scientist indicates the date range of interest.

2. The system analyses the reader comments associated all articles in the specified date range and gives an indication of the level of user interest in the topic based on: the extent of comment on these articles; whether the rate of comment on the most recent articles is greater/lesser than before; whether comments reflect agreement or disagreement amongst commenters; the strength of feeling expressed on the topic and whether this is fairly consistent or highly polarised.

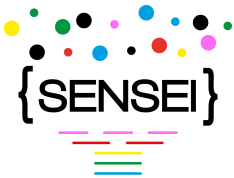### 2.6.6 Use Case 6: Comment Editor Making Content from the Comments

Actor: Comment editor

Goal: to obtain content from the comments for use in an on-line news editorial dedicated to user comments, e.g. The Guardian's, "Comment is Free" page[4]. Specifically, this will include (1), a summary of the top stories of the day (or a given time period) to emerge from the comments, as indicated by volume and strength of comment in different conversations. This might comprise a "headline" summary for each of the top five conversations in relation to the original news article, the headline being a hyperlink to a fuller Town Hall Summary (See Use Case 1) of that conversation. In addition, the goal is to collect (2), a selection of comments from all conversations in a given period to be presented as "editor's picks".

Steps:

1. The comment editor selects a set of conversations and news articles as input to the digest page.

---

[4] http://www.theguardian.com/uk/commentisfree

2. The system analyses the comments and, based on the volume and strength of comments (e.g. in terms of polarity and sentiment and user ratings), identifies the top 5 conversations.

3. The system returns a "headline", i.e. an overarching summary characterising the conversation in relation to the news article for each conversation.

4. The system takes these top ranked conversations as input to use case 1 and generates a town hall summary for each story. These are then linked to the headline summary generated in the previous step.

5. The system also scans the full set of comments in a given time period and uses various algorithms to determine which comments are suitable as "picks". This can be done using a combination of "interestingness" and "authority" measures. Machine learning might also be used to suggest comments, taking the existing editors' picks and the full sets of comments as training data.

6. The system returns a ranked list, showing the top n comments for use in the on-line comment digest page.

7. The system may also periodically scan recent comments and add the best ones to a running "picks" list.

2.6.7 Use Case Priorities

SENSEI needs to prioritize which use cases to focus on in the coming period. We propose to focus first on:

- Use Case 1: Town Hall Summary

- Use Case 3: Back grounding: Looking in Greater Depth at a Comment Poster

- Use Case 5: Identifying Trends in Reader Comments

This will give us one use case from each of the three groups of pairs of use cases we mentioned at the outset of this section, i.e. one use case yielding a digest of the entire set comments related to a single news article, one starting with a single reader comment and building context to help users interpret that post, and one taking multiple conversations (set comments related to a single news article) as inputs and providing meta-level digests or analyses.

2.7 Future Work: Gathering Feedback on Use Cases

As the next step in use case development, we plan to gather feedback data from users on our characterization of the user groups and on our proposed use cases and refine or modify these accordingly. Since users' feedback is likely to be based on the technologies and processes they are already familiar with, ideally we need to include a range of activities to elicit both their current practices and the needs for new technologies that they may not currently be able to articulate. However, our access to busy working journalists will be limited and techniques for eliciting feedback from them need to take this into account.

We propose two main routes for gathering data and feedback from users on our characterization of the user groups and our proposed use cases.

1. Small informal focus groups.
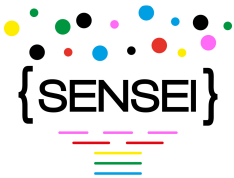
2. One-to-one interviews or questionnaires. We carry out short informal interviews by phone/or in person. We may also use online questionnaires.

Participants and Recruiting:

There are two main groups of participants:

a. Those representing the interests of professional news producers. These include: USFD Journalism students; colleagues in the Department of Journalism Studies; journalists and editors at local and national newspapers, particularly those at *The Independent* and *The Guardian*, who have agreed to collaborate with us. These can be contacted directly by word-of-mouth.

b. Those representing interests of news readers and comment providers. To recruit these we will exploit USFD's "Volunteers" email list which reaches all members of staff and students at the University and possibly advertise directly on newspaper comment sites, if the newspapers are willing.

We propose to use interviews and questionnaires to get feedback from both groups 1 and 2. We also plan to arrange a focus group for group 2; and group 1 (if possible).

Methods:

Record interviews and focus group discussion, if possible. But the main thing will be taking detailed manual notes. And if necessary, we can get people to answer a short, form like set of questions in a written questionnaire.

1. **Focus groups:** Groups of users from selected user categories identified in the previous sections will be invited to focus groups. Focus groups will involve question answering, free conversation, activities that simulate current processes, and engagement with mock, most likely paper-based prototypes in which users will be confronted with novel processes and functionalities outlined in the previous section. If possible we will video record focus groups in order to have access to the full set of information for later analysis. At the same time, observation and note-taking during the session will be applied to record the focus group results. If it proves infeasible to arrange focus groups with any of our user categories, due to, e.g., scheduling issues we shall rely exclusively on one-to-one interviews and questionnaires.

2. **One-to-one interviews and/or online questionnaires:** Interviews and questionnaires will complement or substitute for data collection from focus groups. The specific design of these and the choice of interview versus online questionnaire will depend on our success in establishing focus groups for the user groups. If focus groups are established then the observations from them will be used in designing the interview protocol and/or questionnaires. Otherwise interview protocol and/or questionnaires will be designed to assess the likely utility of our use cases and to reveal functionality that we may have overlooked.

3. Speech and social media scenarios: commonalities and differences

In this section, we identify commonalities and differences between the two application scenarios described in Sections 1 and 2. We divide the discussion under four subheadings, as to whether the remarks relate to: (i) the form of data, (ii) the users, (iii) existing precedents for the tasks that are enabled, and (iv) the functionalities required. We will refer to the social media use case a SM and the speech use case as Speech.

3.1 Form of Data

Similarities:

- Both scenarios generate human conversation data.

- The language in both cases is mostly informal -- colloquial and vernacular -- and not formal.

- Spoken utterance transcription and users' comments are similar in that inter-document boundaries are not present (e.g. punctuations) or noisy. Human conversations or users' do not follow a predefined editorial (e.g. journalistic) norm.

Differences:

- Conversations on social media platforms do not require temporal synchronization (e.g. turn taking) amongst participants, while human spoken conversations do.

- One scenario involves spoken language data, the other written language.

- In addition to the transcript of what has been said, the speech case may be able exploit other aspects of the speech signal, such as intonation and emphasis, which may serve to indicate important content and provides cues to sentiment analysis.

- The conversations in Speech are dialogues, i.e. two party conversations while the SM conversations are multi-party.

- In the SM case the data include an initial article, which serves as a trigger and focus for the conversation; no such trigger, other than the caller's intent, exists in the Speech case.

- In the Speech case the conversation is partly driven by a script (agent's side) on the call centre side, which to some extent determines which topics maybe addressed.
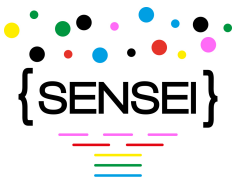
3.2 The Users

Similarities:

- End-user in both scenarios are interested in analytics/summarization over large volumes of conversational data to make their operations more efficient and effective.

Differences:

- In the Speech scenario users are all from different business units of the outsourcing call centre company.

- In the SM case, users include both members of the news production organization (e.g. editors, reporters) and news consumers (e.g. comment posters and readers).

3.3 Precedents for Tasks

- For speech, the intended tasks are essentially the same, or close to, tasks that are already carried out as part of call centre procedure, and even tasks for which some automating software products (albeit ones of limited effectiveness) have already been marketed commercially. Thus, the use cases proposed above are not particularly speculative, since to a large extent they reflect current practice (though the scale on which they are currently carried out is limited due to data volumes and speed). Furthermore, since the use cases relate closely to current practice evaluation data should be relatively easy to acquire.

- For the SM case, this is much less the case, with exceptions to this pattern being the possibility of automating functionality to select 'Editor Picks' comments, and to produce 'What are users saying?' summaries of opinions that are trending within comments. Thus, the use cases introduced above for the SM case for the most part propose novel processes and no readily available evaluation data, in the form of outputs from current practices, exists. Unlike the Speech case, no software products have been marketed to carry out analytics or summarization over on-line reader comment.

3.4 Functionalities

Similarities:

- There are key functionalities that will be required across the two scenarios: particularly for topic analysis (including segmentation, labelling and clustering), conversational summarization, and sentiment analysis.

Differences:

- The SM case requires some functionalities not required for Speech, such as summarization of the non-conversation text of the news article, and topical linkage between conversation and article.

- Functionalities that are common to both scenarios may have implementation differences in the two cases, e.g. such as that sentiment analysis for Speech may be able to exploit cues from the acoustic signal, which are unavailable for SM.

3.5 Conclusion

We have described and compiled a range of Speech Use Case and Social Media Use Case scenarios. Such scenarios include the descriptions of the end-users and functionalities SENSEI's summarization technology may support. The coverage of the use cases has been achieved by preliminary interview with industry partners, SENSEI's affiliates and literature reviews. In the future (second part of Y1), we will a) select the scenarios that will be implemented In SENSEI's project timeframe and b) refine user requirements in the selected scenarios through the use of common methodologies, specifically interviews/questionnaires directed at individuals in specific user groups and the creation of focus groups within user categories where possible.
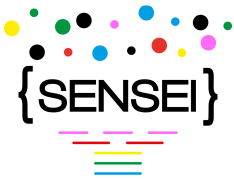
REFERENCES

[Alam & Riccardi 2013] Alam, F., Riccardi, G. (2013). 'Comparative Study of Speaker Personality Traits Recognition in Conversational and Broadcast News Speech', In Proceedings of Interspeech 2013, Lyon

[Ashkanasy 2011] Ashkanasy Neal M. (2011). 'Current emotion research in organizational behaviour'. In Emotion Review, April 2011 vol. 3 no. 2, 214-224

[Bailor 2006] Bailor Coreen (2006) 'The Why Factor in Speech Analytics About'. In Destination CRM. Retrieved 27-04-2014.

[Baldoni et al. 2012]. Baldoni, M., Baroglio, C. Patti, V., and Rena, P. (2012). 'From tags to emotions: Ontology driven semantic analysis in the social semantic web". In Intelligenza Artificiale, Vol. 6, N. 1, 41-54.

[Bender et al. 2011] Bender, Emily M., Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson, Alex Marin, Bin Zhang, and Mari Ostendorf (2011) 'Annotating Social Acts: Authority Claims and Alignment Moves in Wikipedia Talk Pages'. In Proceedings of the Workshop on Languages in Social Media, 48–57. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics.

[Carenini et al. 2007] Carenini, Giuseppe, Raymond T. Ng, and Xiaodong Zhou (2007) 'Summarizing Email Conversations with Clue Words'. In Proceedings of the 16th International Conference on World Wide Web, 91–100. WWW '07. New York, NY, USA.

[Carenini et al., 2012] Carenini, Giuseppe, Jackie Chi Kit Cheung and Adam Pauls. Multi-document summarization of evaluative text. (2012). *Computational Intelligence*.

[Corston-Oliver et al. 2004] Corston-Oliver, Simon, Eric Ringger, Michael Gamon and Richard Campbell (2004) Task-Focused Summarization of Email. In Text Summarization Branches Out: Proceedings of the ACL-04 Workshop , 43--50.

[Diakopoulos and Naaman, 2011] Diakopoulos N., Naaman M., *(2011)* Towards quality discourse in online news comments. In: Hinds PJ, Tang JC, Wang J, et al . (eds) Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work. New York: Association for Computing Machinery, pp. 133–142*.*

[Di Fabbrizio et al., 2011] Di Fabbrizio, Giuseppe, Ahmet Aker and Robert Gaizauskas. STARLET: Multi-document summarization of service and product reviews with balanced rating distributions (2011) In Proceedings of the 2011 IEEE International Conference on Data Mining (ICDM) Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE), Vancouver.

[Di Fabbrizio et al., 2013] G Di Fabbrizio, AJ Stent, R Gaizauskas (2013) Summarizing opinion-related Information for mobile devices. In Neustein, Amy, Markowitz, Judith A. (Eds.) Mobile Speech and Advanced Natural Language.

[Drew 2005] Drew, P. 'Conversation Analysis' (2005) In Handbook of Language and Social Interaction, edited by K.L. Fitch and R.E. Sanders, 71–102. Routledge Communication Series. USA: Ps
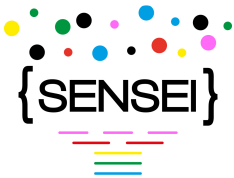
[Hultgren 2011] Hultgren, A. K. (2011), 'Building rapport' with customers across the world: The global diffusion of a call centre speech style. Journal of Sociolinguistics, 15: 36–64.

[Jenkins et al. 2010], Jenkins, S., Delbridge, R., and Roberts, A. (2010). 'Emotional management in a mass customised call centre: examining skill and knowledgeability in interactive service work', In Work Employment & Society, vol. 24 no. 3 546-564

[Liao et al. 2009] Liao, H., Toya, K., Lepak, D. and Hong, Y. (2009). 'Do they see eye to eye? Management and employee perspectives of high-performance work systems and influence processes on service quality'. Journal of Applied Psychology, 94: 2, 371–391.

[Manosevitch and Walker, 2009] Manosevitch, E. and Walker, D. (2009) Reader Comments to Online Opinion Journalism: A Space of Public Deliberation. In 10th International Symposium on Online Journalism.

[Mayfield and Penstein Rosé, 2011] Mayfield, Elijah, and Carolyn Penstein Rosé (2011) 'Recognizing Authority in Dialogue with an Integer Linear Programming Constrained Model'. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies – Volume 1, 1018–1026. HLT '11. Stroudsburg, PA, USA: Association for Computational Linguistics.

[Maynard et al. 2014a] Maynard, D; Bontcheva, K; Rout, D (2014) Challenges in developing opinion mining tools for social media. @NLP can u tag #usergeneratedcontent workshop at LREC 2014.

[Maynard et al. 2014b] Maynard, D; Greenwood, M. Who cares about sarcastic tweets? (2014). Investigating the impact of sarcasm on sentiment analysis. LREC 2014.

[Maynard et al. under review] Maynard, D; Gossen, G; Funk, A. Should I care about your opinion? Detection of opinion interestingness and dynamics in social media. Submitted to Future Internet.

[McKeown and Radev, 1995], McKeown, Kathleen and Dragomir R. Radev (1995) Generating summaries of multiple news articles. *In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval* , 74--82.

[McKeown et al., 2002] McKeown, Kathleen R., Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Judith L. Klavans, Ani Nenkova, Carl Sable, Barry Schiffman and Sergey Sigelman (2002) Tracking and summarizing news on a daily basis with Columbia's Newsblaster. *In Proceedings of the second international conference on Human Language Technology Research* , 280--285.

[Miller, 2004] Miller, K. (2004) Communication Theories: Perspectives, Processes and Contexts. McGraw Hill.

[Murray et al. 2008] Murray, Gabriel and Giuseppe Carenini (2008) Summarizing spoken and written conversations. In Proceedings of the Conference on Empirical Methods in Natural Language Processing , 773--782.

Nenkova and McKeown, 2011] Nenkova, A. and K. McKeown (2011) Automatic Summarization. *Foundations and Trends in Information Retrieval 5(2-3)*, 103-233.

[Nielsen, 2013] Carolyn E Nielsen (2013) "Coproduction or cohabitation: Are anonymous online comment newspaper websites shaping news content?" New Media Society, 7 June 2013.
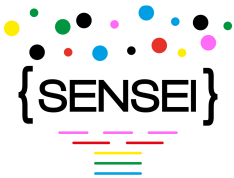
[Pang and Lee 2008] Pang, Bo and Lillian Lee (2008) Opinion Mining and Sentiment Analysis. Found. Trends Inf. Retr. 2 , 1--135.

[Radev et al., 2004] Radev, Dragomir R., Timothy Allison, Sasha Blair-Goldensohn, John Blitzer, Arda Celebi, Stanko Dimitrov, Elliott Drabek, Ali Hakim, Wai Lam, Danyu Liu, Jahna Otterbacher, Hong Qi, Horacio Saggion, Simone Teufel, Michael Topper, Adam Winkel and Zhu Zhang (2004) MEAD - A Platform for Multidocument Multilingual Text Summarization. *In LREC* , 2004.

[Rambow et al. 2004] Rambow, Owen, Lokesh Shrestha, John Chen and Chirsty Lauridsen (2004) Summarizing email threads. In Proceedings of HLT-NAACL 2004: Short Papers , 105--108.

[Renals, 2004]. Renals, Steve (2004) Ami: Augmented multiparty interaction. In Proc. NIST Meeting Transcription Workshop. ICASSP 2004. Montreal. Canada.

[Ruiz et al., 2011] Ruiz, Carlos, David Domingo, Josep Lluís Micó, Javier Díaz-Noci, Koldo Meso and Pere Masip (2011) Public Sphere 2.0? The Democratic Qualities of Citizen Debates in Online Newspapers. *The International Journal of Press/Politics 16(4)* , 463-487.

[Strandberg and Berg 2013] Kim Strandberg, Janne Berg (2013) Online Newspapers' Readers' Comments - Democratic Conversation Platforms or Virtual Soapboxes? Comunicação e Sociedade, vol. 23, 2013, pp. 132 – 152.

[Taboada et al.] Taboada, M; Brooke, J; Tofiloski, M; Voll, K; Stede, M. (2011) Lexicon-based methods for sentiment analysis. Computational Linguistics, 1, 1—41.

[Tamura et al., 2011] Tamura, Akihiro, Kai Ishikawa, Masahiro Saikou and Masaaki Tsuchida (2011) Extractive Summarization Method for Contact Centre Dialogues based on Call Logs. *In Proceedings of 5th International Joint Conference on Natural Language Processing* , 500--508.

[Tucker et al., 2005) Tucker, S., Whittaker, S., & Laban, R. (2005). Identifying user requirements for novel interaction capture. In Symposium 'Annotating and Measuring Meeting Behaviour'. Twente. The Netherlands.

[Uthus and Aha, 2011] Uthus, David C. and David W. Aha (2011) Plans Toward Automated Chat Summarization. *In Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages* , 1--7.

[Zechner, 2002a] Zechner, Klaus (2002) Summarization of Spoken Language - Challenges, Methods, and Prospects. Speech Technology Expert eZine, Issue 6.

[Zechner, 2002b] Zechner, Klaus (2002) Automatic summarization of open-domain multiparty dialogues in diverse genres. *Computational Linguistics 28(4)*, 447--485.

[Zhou and Hovy, 2006] Zhou, Liang and Eduard H. Hovy (2006) On the Summarization of Dynamically Introduced Information: Online Discussions and Blogs. *In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs* , 237-.
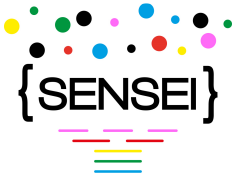
APPENDIX 1

See below a Town Hall Meeting Summary Example.
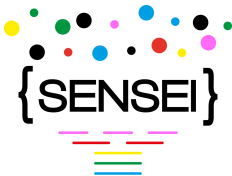
Appendix 1: A Town Hall Meeting Summary Example

A news report of a public meeting is presented in Figure 1:

Householders confront Thames Water boss over floods and sewage

http://www.newburytoday.co.uk/2013/householders-confront-thames-water-boss-over-floods-and-sewage

Thursday, 17th Jan 2013

Reporter: John Garvey Chief Reporter

Email: john.garvey@newburynews.co.uk
Contact: 01635 886628

ANGRY householders, fed up with weeks of raw sewage in their streets and homes, finally confronted a Thames Water boss on Tuesday night.

As frustration built over the flooding and sewage crisis across West Berkshire, the firm tried to defuse public anger by offering a senior manager to field questions.

Nevertheless there were some angry scenes as more than 150 people tried to cram into a public meeting in the Memorial Hall in Lambourn.

Thames Water operarions director Bob Collington then faced a barrage of angry comments and questions in the Memorial Hall in Lambourn on Tuesday, while those unable to squeeze inside strained to hear.

Residents in parts of West Berkshire are suffering almost medieval conditions with raw effluent pouring down streets and into gardens, streams and rivers.

Some parents have sent their children to live with relatives for fear of disease while others are afraid to flush their toilets or drain their sinks because it backs up into their homes.

Mr Collington told the meeting, also attended by floods minister Richard Benyon and representatives from West Berkshire Council and the Environment Agency: "I realise this is such a big issue for you. I want to give you reassurance from the highest level that we're trying to help. But we've been overwhelmed throughout the region."

He faced heckling from the audience who reminded him that this was not the first year their homes had been flooded with sewage and that even last summer they had faced similar problems.

Mr Collington was jeered by some as he conceded: "There's no quick fix – no silver bullet."

He was challenged over Thames Water's failure to upgrade the sewage system and one resident said: "Your charges increased on average 6.7 per cent – well above inflation; you had interim profits of £7.6 million and paid share holders a £7.4 million dividend, so the money's there, it's not rocket science."

Mr Collington said that vast amounts were recycled into the company's capital investment programme.

Thames Water spokesman Simon Evans explained later: "We do pay dividends to our shareholders because they finance our activities. However we're currently spending £1 billion a year investing in improvement to our networks."

Another resident asked Mr Collington, to huge applause: "When are you going to invest in a decent sewage system? We can all go home happy if you tell us. No more hand wringing – it's investment we want."

He responded: "I can't stand here and give you a guarantee we're going to upgrade the sewage system in this area. "But we are going to find out what the problem is and find a solution that works in a joint effort with the local authority and Environment Agency."

However this failed to satisfy many people, who claimed they had heard simuilar assurances in years past.

Residents did not go home entirely empty handed, however.

Repeatedly pressed to give a time table for an investigation into possible causes and solutions to the problem, Mr Collington promised to report back within six months.

If we examine the article we can see examples of how the kind of content we describe in section 4 above is used in a news report of the meeting:

- **What the meeting was about and/or why it was called:**
  "The flooding and sewage crisis across West Berkshire"

- **Headline:**
  "Householders confront Thames Water boss over floods and sewage"

- **Background to the meeting:**
  "Residents in parts of West Berkshire are suffering almost medieval conditions with raw effluent pouring down streets and into gardens, streams and rivers".
  "Some parents have sent their children to live with relatives for fear of disease while others are afraid to flush their toilets or drain their sinks because it backs up into their homes".
  "As frustration built over the flooding and sewage crisis across West Berkshire, the firm tried to defuse public anger by offering a senior manager to field questions"

- **Where the meeting was held:**
  "Memorial Hall in Lambourn".

- **Who attended:**
  "householders", "residents", "Thames Water" executives/representatives, "Thames Water operations director Bob Collington", "floods minister Richard Benyon and representatives from West Berkshire Council and the Environment Agency".

- **How many attended:**
  "150 people tried to cram into a public meeting in the Memorial Hall in Lambourn."

- **Who were the main speakers:**
  "Thames Water operations director Bob Collington", "floods minister Richard Benyon and representatives from West Berkshire Council and the Environment Agency"

- **Who was in the audience:**
  "Householders" and "residents"…

- **Main issue raised/ most of the discussion focused on […]**
  The problem of sewage in West Berkshire .

- **Key issues were […] and who raised them and who they were put to**
  "He [Mr Collington] was challenged over Thames Water's failure to upgrade the sewage system"
  "[Mr Collington was] Repeatedly pressed to give a time table for an investigation into possible causes and solutions to the problem"
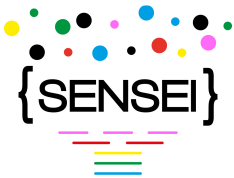
- **There was strong support, by/of whom, for issues […] :**
"Another resident asked Mr Collington, **to huge applause**: "When are you going to invest in a decent sewage system? We can all go home happy if you tell us. No more hand wringing – it's investment we want.""

- **There was strong criticism, by/of whom, for issues […]:**
"Thames Water operations director Bob Collington then **faced a barrage of angry comments** and questions in the Memorial Hall in Lambourn on Tuesday"
Mr Collington was jeered by some as he conceded: "There's no quick fix – no silver bullet.""

- **Selected quotes to illustrate opinion on key issues:**
"one resident said: "Your charges increased on average 6.7 per cent – well above inflation; you had interim profits of £7.6 million and paid share holders a £7.4 million dividend, so the money's there, it's not rocket science.""

- **The most striking outcomes /the most widely held opinion:**
"Repeatedly pressed to give a time table for an investigation into possible causes and solutions to the problem, **Mr Collington promised to report back within six months**".

"**there were some angry scenes** as more than 150 people tried to cram into a public meeting in the Memorial Hall in Lambourn".