

## **D7.4 – Interim Report on Dissemination, Exploitation and Intellectual Property**

Document Number	D7.4
Document Title	Interim Report on Dissemination, Exploitation and Intellectual Property
Version	3.5
Status	Final
Work Package	WP7
Deliverable Type	Report
Contractual Date of Delivery	31.10.2015
Actual Date of Delivery	30.10.2015
Responsible Unit	UNITN
Keyword List	dissemination activities, exploitation plan, intellectual property
Dissemination level	PU



## **Editor**

Giuseppe Riccardi (University of Trento, UNITN)

## **Contributors**

Elisa Chiarani	(University of Trento, UNITN)
Carmelo Ferrante	(University of Trento, UNITN)
Benoit Favre	(University of Marseille, AMU)
Emina Kurtic	(University of Sheffield, USFD)
M. Kabadjov	(University of Essex, UESSEX)
Hugo Zaragoza	(Websays SL, Websays)
Marc Poch	(Websays SL, Websays)
Vincenzo Giliberti	(Teleperformance Italy, TP)
Letizia Molinari	(Teleperformance Italy, TP)

## **SENSEI Coordinator**

Prof. Giuseppe Riccardi

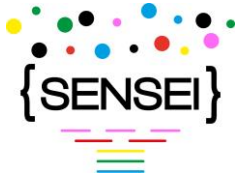
Department of Information Engineering and Computer Science

University of Trento, Italy

[giuseppe.riccardi@unitn.it](mailto:giuseppe.riccardi@unitn.it)

## Document change record

Version	Date	Status	Author (Unit)	Description
0.1	07.07.2015	Draft	E. Chiarani, G. Riccardi (UNITN)	Outline
0.2	19.08.2015	Draft	Marc Poch (Websays)	Websays contribution for dissemination and exploitation activities
0.3	20.08.2015	Draft	Emina Kurtic (USFD)	USFD inputs added
0.4	24.08.2015	Draft	V. Giliberti (TP)	TP inputs provided for dissemination and exploitation activities
0.5	28.08.2015	Draft	M. Kabadjov (UESSEX)	UESSEX contribution
0.6	31.08.2015	Draft	B. Favre (AMU)	AMU info provided
1.0	31.08.2015	Draft	E. Chiarani (UNITN)	Merged initial draft
1.1	10.09.2015	Draft	C. Ferrante (UNITN)	Update of website description and statistics added
1.2	22.09.2015	Draft	H. Zaragoza, M. Poch (Websays)	Exploitation plan Websays added
1.3	25.09.2015	Draft	L. Molinari (TP)	3.1.3 completed
1.4	28.09.2015	Draft	V. Giliberti (TP)	TP contribution for exploitation plan and revision of some inputs
1.5	29/09/2015	Draft	E. Chiarani (UNITN)	Consolidated draft
2.0	02/10/2015	Draft	V. Giliberti (TP), M. Poch (Websays), E. Chiarani (UNITN)	Updated version
3.0	04/10/2015	Draft	Giuseppe Riccardi (UNITN)	Complete Revision and section added
3.1	08/10/2015	Draft	Hugo Zaragoza (Websays)	Review
3.2	16/10/2015	Draft	Elisa Chiarani (UNITN)	Review comments addressed. Updated version
3.3	20/10/2015	Draft	Vincenzo Giliberti (TP)	Updated version
3.4	20/10/2015	Draft	Giuseppe Riccardi (UNITN)	Revised version
3.5	28/10/2015	Final	Elisa Chiarani (UNITN)	Final version for submission



## Executive Summary

The document reports on the dissemination, exploitation and intellectual property (IP) activities accomplished during the second Period of SENSEI. In this deliverable we report on the dissemination products we have delivered and events we have organized to reach both the scientific and technology community as well as the end-users. The last part of the document includes the achievements and plans for the exploitation of SENSEI outputs.

# Table of Content

EXECUTIVE SUMMARY .....	4
1. INTRODUCTION .....	6
1.1 FOLLOW UP TO RECOMMENDATIONS FROM THE FIRST REVIEW .....	6
2. PROJECT WEBSITE .....	8
3. DISSEMINATION AND COMMUNICATION.....	11
3.1 DATA DISSEMINATION .....	11
3.1.1 Shared Task .....	11
3.2 DISSEMINATION ACTIVITIES .....	13
3.2.1 Conferences, Workshops, Presentations, Events.....	13
3.2.2 Publications Period 2.....	16
3.2.3 Press releases and articles .....	21
3.2.4 Contacts and Cooperation .....	23
3.3 DISSEMINATION PLAN FOR PERIOD 3 .....	26
4. EXPLOITATION ACTIVITIES .....	27
4.1 EXPLOITATION ACTIVITIES PERIOD 2.....	27
4.1.1 Internal Exploitation.....	27
4.1.2 External Exploitation Events.....	28
4.2 EXPLOITATION PLAN .....	32
4.2.1 Conversation Analytics.....	33
4.2.2 From Social Media to NewsPaper companies .....	37
5. MANAGING INTELLECTUAL PROPERTY .....	41
6. CONCLUSIONS .....	42
REFERENCES.....	43

# 1.Introduction

This document reports on the progress of the a) dissemination activities b) cooperation with external entities c) exploitation plans and d) management of intellectual property.

In the following section, we briefly describe how the recommendations related to WP7 have been addressed. In Section 2 we present Period 2 achievements related to task 7.1 namely the web site population. In Section 3, we outline the dissemination events and activities that the SENSEI consortium has participated to or organized (Task 7.2). In Section 4 we outline the exploitation activities carried out. Furthermore, we draft an exploitation plan, especially for the industrial partners, as well as the steps taken towards the implementation of these strategies (Task 7.3). In Section 5, we briefly review issues related to intellectual property management (Task 7.4).

## 1.1 Follow up to Recommendations from the First Review

- **Recommendation n.6:** *The annotated corpora should be checked for entailment, and it should be discussed if and how the tools from the Excitement project, specifically for entailment, can be utilized in the summarization modules.*

SENSEI partners reviewed Excitement entailment platform and met with their team. We have interacted with Excitement members and learned on how to use Excitement tools within SENSEI. UNITN, USFD and AMU have performed experiments with tools on SENSEI's data however the results of these experiments highlighted that at the current state of development and performance, Excitement's tools are not fit for SENSEI's real-life data and/or problems. More details on this in section 3.2.4.1.

- **Recommendation n.9:** *There should be an exploitation plan before the end of the project. It would be advisable to start this work in Y2. This may include a market study and the identification of how parts of the technology will be used in immediate exploitation.*

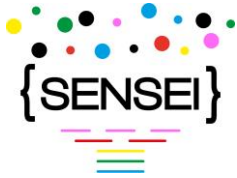
The industrial partners have drafted an exploitation plan for the social media use case and the speech use case. The exploitation plan will be extended and refined during Period 3 to have it ready before the end of the project.

*The academic partners also need to state how they will plan to continue to work with the technology and how they will plan to share the tools developed in this project and maintain them after the end of the project.*

The academic partners agree to make the tools public and open source and to use them after the end of the project in related research activities.

*The ownership of the intellectual property should also be dealt with before the exploitation can start.*

The IP issue is regulated by the Consortium agreement and the concerned partners are dealing coherently with this issue.



- **Recommendation n.10:** *As the consortium partners have been successful in reaching out to the Italian and Spanish press, it would be recommendable to try to reach out to the press in other countries too.*

The concerned partners have dealt with this issue. USFD is in the process of issuing a press release through Sheffield University, two of whose media officers are to be involved in the evaluation stage. That should be ready during Period 3. USFD has a close and continuing working relationship with the UK press through our partnership with 'The Guardian' to release a story as soon as we jointly agree an optimal timing in relation to project progress.

Teleperformance Italy is in touch with Teleperformance France to develop contacts with French media to reach out to the press during Period 3.

## 2. Project website

SENSEI's website has been developed using WordPress and registered under the domain <http://www.sensei-conversation.eu> in accordance with the EU recommendation.

The website purpose is twofold. The main purpose is to provide information about the SENSEI project and demos to viewers. The other purpose is to internally share, store and exchange files between the partners involved in the project. Hence, the website platform includes a web interface, which is optimized for mobile, and a back office, which is accessible only to the partners of the SENSEI project.

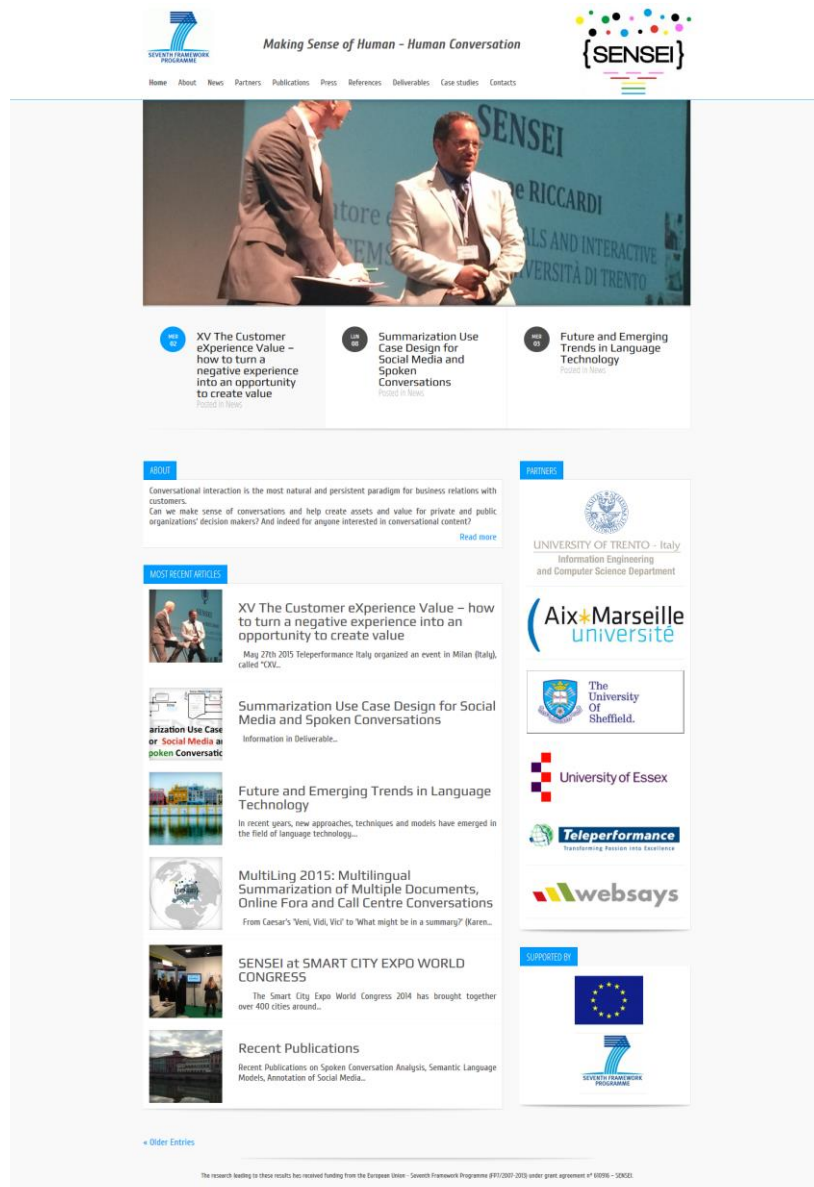


Figure 1 - SENSEI website home page

The website has been updated, by adding news about events and dissemination activities that took place in Period 2, publications and more. Deliverables submitted in Period 1 have been released



through the website; they can be downloaded in pdf format. From the same page it is possible to download the public compressed archive of the deliverable 2.1 containing sample data. Deliverables due in P2 will be uploaded after the official acceptance by EC.

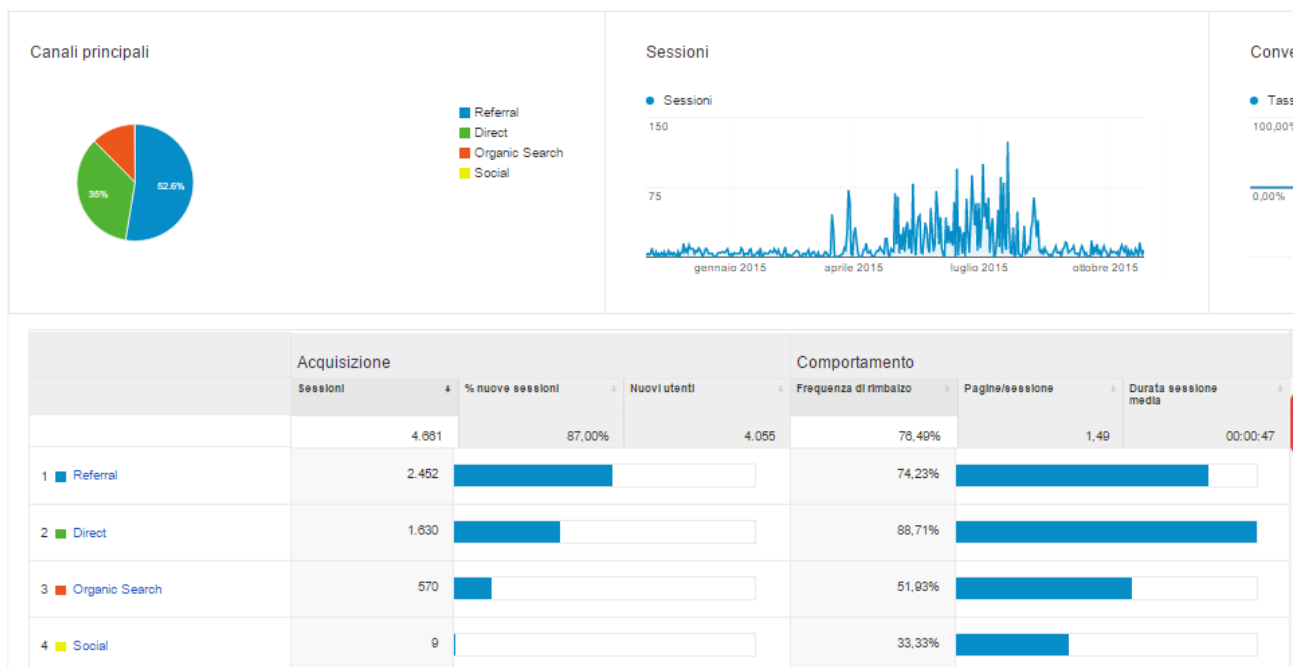
Flyers and posters have been published in low resolution on the page “about” so that they can be freely seen by interested people.

Project’s general contact email is freely visible so that anyone can write us to ask for information and updates.

Recently a statistics plugin have been installed to monitor visits and events on the website to check the interest in the internet world about the project and the different contents of the website.

Since 1<sup>st</sup> November 2014<sup>1</sup> we have had:

- 4069 new visits
- 47 seconds is the average duration of a session
- 1 page and half the average of pages visited in a session
- 1503 visits come from US
- 2452 visits from links in other websites (referral)
- 1630 visits from direct address box
- 570 organic visits from research motors



<sup>1</sup> Last updated on 29 October 2015

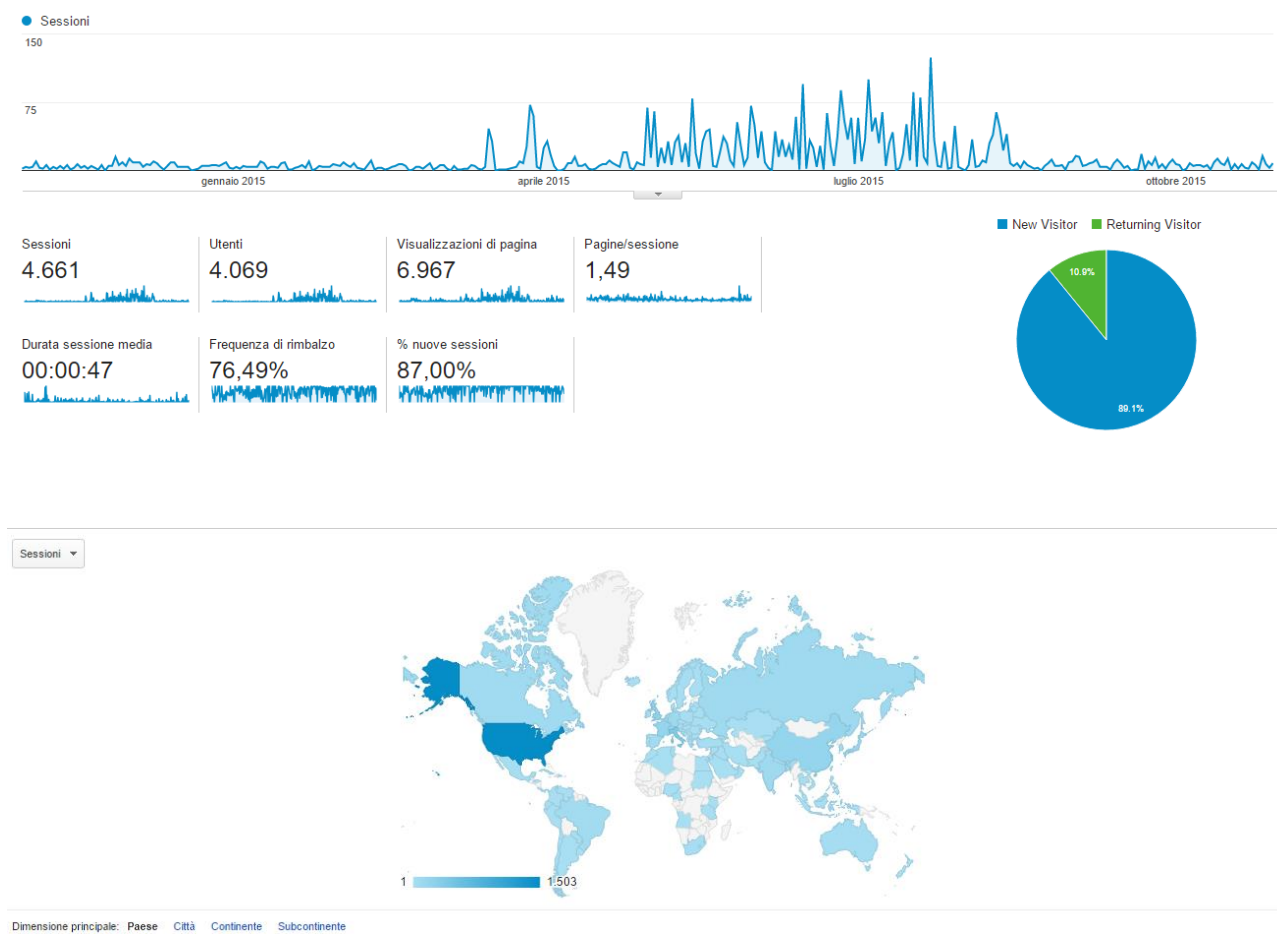


Figure 2: SENSEI Website statistics 1 November 2014 – 29 October 2015

## 3. Dissemination and communication

### 3.1 Data dissemination

#### 3.1.1 Shared Task

The SENSEI consortium organized two shared tasks at the Multiling 2015 multilingual summarization evaluation campaign: the Online Forum Summarization (OnForumS) task and the Call-Centre Conversation Summarization (CCCS) task, along with the traditional tasks of the evaluation, Wikipedia article summarization and news article summarization. The results of the evaluation have been presented at the SIGDIAL conference in Prague, September 1-4, 2015. For the data distribution guidelines we followed the guidelines outlined in D8.4.

The OnForumS task investigates how the mass of comments found on news providers' web sites can be summarized. We posit that a crucial initial step towards that goal is to determine which comments link to, be that either specific points within the text of the article, the global topic of the article, or comments made by other users. This constitutes a linking task. Furthermore, a set of types or labels for a given link may be articulated to capture phenomena such as agreement (e.g., in favour, against) and sentiment (e.g., positive or negative) with respect to the comment target. As such, the OnForumS task is a pioneering attempt at encompassing automatic summarization, argumentation mining and sentiment analysis into one shared task and at bringing crowdsourcing to the evaluation of systems for automatic summarization and argument structure parsing.

Data for the task was collected in English and Italian. A sample data consisting of one article in English and small set of comments and labelled links result of internal pre-pilots was released early on. Four research groups (Univ. of Posts and Telecommunications of Beijing, Joint Research Centre of the European Commission, Univ. Sheffield and Univ. of West Bohemia), each submitting two runs, participated in the task and these complemented with two baseline system runs were evaluated via crowdsourcing. Results and evaluation protocol are detailed in "OnForumS: A Shared Task on On-line Forum Summarisation", M. Kabadjov et al. (task overview report), MultiLing 2015.

The CCCS task consists in automatically generating summaries of spoken conversations in the form of textual synopses that shall inform on the content of a conversation and might be used for browsing a large database of recordings. Compared to news summarization where extractive approaches have been very successful, the CCCS task's objective was to foster work on abstractive summarization in order to depict what happened in a conversation instead of what people actually said.

The track leverages conversations from the Decoda and Luna corpora of French and Italian call centre recordings, both with transcripts available in their original language as well as English translation (both manual and automatic) provided by SENSEI. In the public transportation and help desk domains, the dialogs offer a rich range of situations (with emotions such as anger or frustration) while staying in a coherent and focused domain.

Given transcripts, participants to the CCCS task had to generate abstractive summaries (called synopses in SENSEI) informing a reader about the main events of the conversations, such as the objective of the caller, whether and how it was solved by the agent, and the attitude of both parties.



Evaluation has been performed by comparing submissions to reference synopses written by quality assurance experts from call centres.

Two groups (Univ. Avignon and Univ. Taiwan) submitted four systems to the CCCS task, which were confronted to three extractive baselines, and evaluated against reference synopses with the ROUGE evaluation metric. Results and evaluation protocol are detailed in “Call Centre Conversation Summarization: A Pilot Task at Multiling 2015”, Benoit Favre et al., SIGDIAL 2015.

The SIGDIAL presentations were an opportunity to meet with the summarization community which gave good feedback and willingness to continue organizing summarization evaluation campaigns, in particular reiterate both of the SENSEI tasks, and develop the community at the European level for instance with a COST. For the SENSEI project, the organization of the shared task has been an opportunity to reach out to the community and create a well-defined evaluation benchmark on tasks that matter for the project, and for the community it has been an opening towards new research directions for which the SENSEI project lowered the barrier of entry.

## 3.2 Dissemination activities

### 3.2.1 Conferences, Workshops, Presentations, Events

Table 1 shows the Conferences, Workshops and events in which SENSEI project's work and objectives were disseminated through papers and presentations during Period 2.

Table 1 – Conferences and Workshops Period 2

Type of activity	Main Leader (Partner)	Title / Name of the event	Date	Place	Type of Audience: Scientific community	Size of Audience	Countries addressed	Purpose / Justification / Outcomes	Name of the attendees/speaker
Congress	Websays	Smart City Expo World Congress 2014	1/11/2014	Barcelona, Spain	cities, enterprises and entrepreneurs, research centres, universities and other public or non-governmental organizations	10 000	International	Show SENSEI to a large and varied community	Hugo Zaragoza
Workshop presentation	UniTN	ACM Multimedia 2014	7/11/2014	Orlando, FL	Scientific Community	20-30	international	Presentation of Workshop "The Workshop on Computational Personality Recognition 2014"	Fabio Celli
Workshop	UniTN	SLT 2014	7/12/2014	South Lake Tahoe, USA	Scientific Community	60-100	International	Presentation of the poster "Sematic Language Models for Automatic Speech Recognition" related to WP3: Semantic Parsing	Giuseppe Riccardi, Ali Orkan Bayer
Conference presentation	UniTN	CLiC-it 2014	10/12/2014	Pisa, Italy	Scientific Community Computational Linguistics	100	Italy	Presentation of the paper "Annotation of Complex Emotions in Real-Life Dialogues: The Case of Empathy"	Morena Danieli
Conference presentation	UniTN	CLIC-IT 2014	10/12/2014	Pisa, Italy	Scientific Community, Computational	50-60	Italy	Presentation of the paper "CorEA: Italian News Corpus with Emotions and Agreement"	Fabio Celli, Giuseppe Riccardi

Type of activity	Main Leader (Partner)	Title / Name of the event	Date	Place	Type of Audience: Scientific community	Size of Audience	Countries addressed	Purpose / Justification / Outcomes	Name of the attendees/speaker
					Linguistics				
workshop	Websays	Listening to Social Networks 2015	24/3/2015	Barcelona, Spain	politicians and city councils, universities, technology innovators and integrators	100	Spain	Explain SENSEI goals to politicians and city councils technicians	Hugo Zaragoza, Websays Team
Teaching	AMU	EJCM	25/3/2015	Marseille (FR)	Journalism students	10	France	Presentation of SENSEI project to journalism students	Benoit Favre, Frederic Bechet
Conference presentation	AMU	Joint workshop ACL-ISO on interoperable semantic annotation	1/4/2015	London (GB)	Scientific Community	30-50	International	Presentation of paper "Rapid FrameNet annotation of spoken conversation transcripts" related to WP3: Semantic parsing	Jeremy Trione
Conference presentation	UNITN	ICASSP 2015	21/04/2015	Brisbane, Australia	Scientific Community	60-100	International	Presentation of the poster "ANNOTATING AND CATEGORIZING COMPETITION IN OVERLAP SPEECH" related to WP3: Extraction of non-verbal features for para-semantic prediction	Giuseppe Riccardi
Workshop	TP	CXV The Customer eXperience Value - how to turn a negative experience into an opportunity to create value	27/05/2015	Milan, Italy	Clients in several industries	50	Italy	Explain SENSEI goals to C-levels from several industries	Giuseppe Riccardi (UNITN), Gabriele Piva (Ceo TP) Paolo Righetti (Ceo GN)
Conference presentation	AMU	11ème Défi Fouille de Texte (DEFT'2015) (DEFT 2015)	22/6/2015	Caen (FR)	Scientific community	30-50	France	Presentation of SENSEI participation to DEFT sentiment analysis evaluation campaign (paper "TALEP @ DEFT'15 : Le plus coool des systèmes d'analyse de sentiment") related to WP3:	Mickael Rouvier

Type of activity	Main Leader (Partner)	Title / Name of the event	Date	Place	Type of Audience: Scientific community	Size of Audience	Countries addressed	Purpose / Justification / Outcomes	Name of the attendees/speaker
								Parasemantic, Adaptation	
Exhibition	AMU	Salon de l'innovation / TALN	23-25/06/2015	Caen (FR)	Scientific / business	150	France	Presentation of SENSEI project to businesses and scientists at TALN conference	Frederic Bechet
Conference presentation	Websays	The future of Smart City platforms: from sensors to public opinion and sentiment analysis.	24/06/2015	Rome, Italy	politicians and city councils, universities, technology innovators and integrators	100	International	Explain SENSEI goals to politicians and city councils technicians	Hugo Zaragoza (Websays)
TedX talk	Websays	TedX IE Madrid	09/07/2015	Madrid, Spain	Students, citizens, academics	100	Spain	In the talk "Our Opinions: A New Raw Material" the project SENSEI is presented: <a href="https://youtu.be/uRgqrkVrNqI">https://youtu.be/uRgqrkVrNqI</a>	Hugo Zaragoza (Websays)
Conference presentation	UNITN	CoNLL 2015	30/07/2015	Beijing, China	Scientific Community	60-100	International	Presentation of the paper "The UniTN Discourse Parser in CoNLL 2015 Shared Task: Token-level Sequence Labeling with Argument-specific Models" related to WP4: Discourse Parsing	Evgeny A. Stepanov (UNITN)
Conference presentation	USFD	MultiLing OnForumS shared task @ SigDial 2015	1/09/2015	Prague, Czech Republic	Scientific community	30-50	International	Presentation of the paper "Sheffield-Trento System for Sentiment and Argument Structure Enhanced Comment-to-Article Linking in the Online News Domain" related to WP5: linking and summarizing news article comments	Adam Funk (TBC)
Conference presentation	USFD	SigDial 2015	1/09/2015	Prague, Czech Republic	Scientific community	30-50	International	Presentation of the paper "Comment-to-Article Linking in the Online News Domain" related to WP5: linking and summarizing news article comments	Adam Funk (TBC)
Conference presentation	AMU	SigDial 2015	3/09/2015	Prague (CZ)	Scientific Community	30-50	International	Presentation of paper "Call Centre Conversation Summarization: A	Benoit Favre (AMU)

Type of activity	Main Leader (Partner)	Title / Name of the event	Date	Place	Type of Audience: Scientific community	Size of Audience	Countries addressed	Purpose / Justification / Outcomes	Name of the attendees/speaker
								Pilot Task at Multiling 2015"	
Conference presentation	UNiTN	Interspeech 2015	6/09/2015	Dresden, Germany	Scientific Community	60-100	International	Presentation of the poster "Deep Semantic Encodings for Language Modeling" related to WP3: Semantic Parsing	Giuseppe Riccardi (UNiTN), Ali Orkan Bayer (UNiTN)
Conference presentation	UNiTN	Interspeech 2015	06/9/2015	Dresden (DE)	Scientific Community	60-100	International	Presentation of the paper: "The Role of Speakers and Context in Classifying Competition in Overlapping Speech" related to WP3: Extraction of non-verbal features for para-semantic prediction	Giuseppe Riccardi (UNiTN), Shammur Absar (UNiTN)
Conference presentation	AMU	Interspeech 2015	8/09/2015	Dresden (DE)	Scientific Community	50-150	International	Presentation of the paper "Adapting Lexical Representation and OOV Handling from Written to Spoken Language with Word Embedding" related to WP3: Adaptation	Jeremie Tafforeau, Frederic Bechet
Workshop in conference	UNiTN	CLEF2015	8-11/9/2015	Toulouse, France	Scientific community	50-70	International	presentation of "Overview of the 3rd Author Profiling Task at PAN 2015"	Fabio Celli (UNiTN)

### 3.2.2 Publications Period 2

All the publications are open access and can be downloaded in pdf-format from the SENSEI website.

Details of the publications funded by the project have been uploaded to the agreed Bibliographic social network <http://www.mendeley.com/> and tagged with the tag "SENSEI- 610916".

Table 2 – Publications Period 2



Title of the article	Name of Authors and Organisation	Publication	Relevant pages	Year of Publication	Permanent identifiers	Can be open access provided to this publication?	Relevance for SENSEI
The Workshop on Computational Personality Recognition 2014	F. Celli, B. Lepri, D. Gatica-Perez, G. Riccardi, F. Pianesi	Proceeding MM '14 Proceedings of the ACM International Conference on Multimedia	1245-1246	2014	DOI: 10.1145/2647868.2647870	Yes	Related to WP3 Feature extraction
Semantic language models for Automatic Speech Recognition	A.O. Bayer, G. Riccardi	Proceedings of Spoken Language Technology Workshop (SLT), 2014 IEEE	7-12	2014	DOI: 10.1109/SLT.2014.7078541	Yes	Related to WP3: Semantic Parsing
Annotation of Complex Emotions in Real-Life Dialogues: The Case of Empathy	M. Danieli, G. Riccardi and F. Alam	Proceedings of CLIC-IT 2014	122-127	2014		Yes	Related to WP5
Presentation of the paper CorEA: Italian News Corpus with Emotions and Agreement	F. Calli, G. Riccardi	Proceedings of CLIC-IT 2014	98-102	2014		Yes	Related to WP2 Data collection , WP3 Annotation
Rapid FrameNet annotation of spoken conversation transcripts	Jeremy Trione, Frederic Bechet, Benoit Favre, Alexis Nasr	In Proceedings 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)	-	2015		Yes	Related to WP3: Semantic parsing
Annotating and categorizing competition in overlap speech	Morena Danieli, Shammur Absar Chowdhury, Giuseppe Riccardi	Proceedings of ICASSP 2015	5316 - 5320	2015	DOI: 10.1109/ICASSP.2015.7178986	Yes	related to WP3: Extraction of non-verbal features for para-semantic prediction
TALEP @ DEFT'15 : Le plus coool des systèmes d'analyse de sentiment	Mickael Rouvier, Benoit Favre, Balamurali Andiyakkal Rajendran	Actes de la 11e DÃ©fi Fouiille de Texte	97-103	2015		Yes	related to WP3: Parasemantic, Adaptation

Title of the article	Name of Authors and Organisation	Publication	Relevant pages	Year of Publication	Permanent identifiers	Can be open access provided to this publication?	Relevance for SENSEI
The UniTN Discourse Parser in CoNLL 2015 Shared Task: Token-level Sequence Labeling with Argument-specific Models	Evgeny Stepanov; Giuseppe Riccardi; Ali Orkan Bayer	Proceedings of the Nineteenth Conference on Computational Natural Language Learning (CoNLL 2015)	25-31	2015		Yes	related to WP4: Discourse Parsing
Sheffield-Trento System for Sentiment and Argument Structure Enhanced Comment-to-Article Linking in the Online News Domain	Aker, A., Celli, F., Funk, A., Kurtic, E., Hepple, M., Gaizauskas, R.	In Proceedings of MultiLing2015, in conjunction with SigDial 2015	-				related to WP5: linking and summarizing news article comments
MultiLing 2015: Multilingual Summarization of Single and Multi-Documents, On-line Fora, and Call-center Conversations	G. Giannakopoulos, J. Kubina, J. Conroy, J. Steinberger, B. Favre, M. Kabadjov, U. Kruschwitz, M. Poesio	In Proceedings of SIGDIAL 2015	270–274	2015		Yes	Related to WP4
OnForumS: A Shared Task on On-line Forum Summarisation	M. Kabadjov, J. Steinberger, U. Kruschwitz, M. Poesio	In Proceedings Addendum of SIGdial/MultiLing Special Session (On-line: <a href="http://multiling.iit.demokritos.gr/pages/view/1573/multiling-2015-proceedings-addendum">http://multiling.iit.demokritos.gr/pages/view/1573/multiling-2015-proceedings-addendum</a> )	232-236	2015		Yes	Related to WP4
Call Centre Conversation Summarization: A Pilot Task at Multiling 2015	Benoit Favre, Evgeny Stepanov, Jeremy Trione, Frederic Bechet and Giuseppe Riccardi	Proceedings of the SIGDIAL 2015 Conference	232-236	2015		Yes	Related to WP5

Title of the article	Name of Authors and Organisation	Publication	Relevant pages	Year of Publication	Permanent identifiers	Can be open access provided to this publication?	Relevance for SENSEI
Adapting lexical representation and OOV handling from written to spoken language with word embedding	Jeremie Tafforeau, Thierry Artieres, Benoit Favre, Frederic Bechet	Proceedings of Interspeech 2015		2015		Yes	related to WP3: Adaptation
Comment-to-Article Linking in the Online News Domain	Ahmet Aker, Fabio Celli*, Adam Funk, Emina Kurtic, Mark Hepple, Rob Gaizauskas	Proceedings of the SIGDIAL 2015 Conference	245-249	2015		Yes	related to WP5: linking and summarizing news article comments
Deep Semantic Encodings for Language Modeling	Ali Orkan Bayer, Giuseppe Riccardi	Proceedings of Interspeech 2015		2015		Yes	related to WP3: Semantic Parsing
The Role of Speakers and Context in Classifying Competition in Overlapping Speech	Morena Danieli, Shammur Absar Chowdhury, Giuseppe Riccardi	Proceedings of Interspeech 2015	1844-1848	2015		Yes	related to WP3 and WP5
Overview of the 3rd Author Profiling Task at PAN 2015	Francisco Manuel Rangel Pardo, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein, Walter Daelemans	Proceedings of CLEF 2015		2015		Yes	WP3: Extraction of non-verbal features for para-semantic prediction
In the mood for sharing content: Emotions, persoality and interaction styles in the diffusion of news.	F. Celli, A. Ghosh, F. Alam, G. Riccardi	Information, Processing and Mangement, Special Issue on emotions and sentiment	tbd	2015		No	Related to WP3 Cross media, feature extraction

Title of the article	Name of Authors and Organisation	Publication	Relevant pages	Year of Publication	Permanent identifiers	Can be open access provided to this publication?	Relevance for SENSEI
OnForumS: the Shared task on Online Forum Summarisation at MultiLing 2015	M. Kabadjov, J. Steinberger, E. Barker, U. Kruschwitz, M. Poesio.	in Proceedings of the 7th Forum for Information Retrieval and Evaluation (FIRE 2015)	tbd	2015		Yes	Related to WP4



### 3.2.3 Press releases and articles

USFD is in the process of issuing a press release through Sheffield University, two of whose media officers are to be involved in the evaluation stage. That should be ready for publication in Period 3. USFD has a close and continuing working relationship with the UK press through our partnership with 'The Guardian'. We are confident that our SENSEI story will be told soon to an audience of newspaper and journal readers, although we would like to reward 'The Guardian' for its cooperation by giving it a special status in our media relations (i.e. priority in publishing about SENSEI). We will work with 'The Guardian' to release a story as soon as we jointly agree an optimal timing in relation to project progress.

Teleperformance Italy is in touch with Teleperformance France to develop contacts with French media to reach out to the press during Period 3.

On October 27<sup>th</sup> 2014. Teleperformance Italy has sent over two thousand employees and customer a newsletter on Innovation Projects and SENSEI.



#### INNOVARE PER VINCERE



Abbiamo scritto altre volte su queste pagine dell'innovazione possibile in TP Italia ma ora possiamo affermare che *"alea iacta est"* ("Il dado è tratto"-Giulio Cesare, circa duemila anni fa), il nostro *"Rubicone"* lo abbiamo passato e indietro non possiamo tornare.

Con l'innovazione TP Italia vince in Europa, come abbiamo fatto superando una concorrenza di 80 competitori nell'ambito del settimo programma quadro ICT dell'Unione Europea, con il **Progetto di ricerca e sviluppo Sensei** sulla prototipazione di un tool per l'analisi delle interazioni telefoniche e web, che stiamo portando avanti con le Università di Trento, Aix-Marsiglia, Sheffield ed Essex.

Con l'innovazione TP Italia vince in Italia, come abbiamo fatto aggiudicandoci una piccola ma strategica commessa, sulla **Speech Analysis**, ovvero sull'analisi dettagliata dei contenuti delle conversazioni lato Clienti, per comprendere dalla loro diretta voce perché stanno contattando il contact center, quali sono i reali bisogni e le azioni che si possono intraprendere per rispondere adeguatamente a tali esigenze.

Con l'innovazione TP Italia può vincere nelle sfide con i nostri competitori, proponendo ad esempio come stiamo facendo, le soluzioni:

- **TPClient** che consente la gestione dei contatti provenienti da differenti canali di interazione e in particolare dai social media come Facebook e Twitter, e il miglioramento della gestione delle interazioni con i Clienti, elaborando un flusso di lavoro automatizzato che segue i casi dei Clienti, segnala quando ha bisogno di essere completato e monitora il tutto fino alla completa risoluzione.
- **Click to Call** che permette ai Clienti che stanno navigando su un sito web di un'azienda, di inserire il proprio numero di telefono ed esser immediatamente richiamato da un operatore del nostro call center per aver supporto nell'acquisto del prodotto o servizio.

Con l'innovazione TP Italia può vincere la sfida della sostenibilità, come stiamo facendo implementando ad esempio soluzioni per il **risparmio energetico** e il monitoraggio pro-attivo della continuità elettrica dei nostri Siti produttivi, ma anche sedendoci da coprotagonisti nei tavoli dove si discute e si decide sulla redazione dei **bilanci sociali di impresa**. L'innovazione vince solo se riesce a pervadere tutta TP Italia e a tal fine abbiamo intrapreso degli **innovation dissemination meeting**, coinvolgendo diverse aree aziendali in workshop tematici bidirezionali nei quali illustriamo queste innovazioni e raccogliamo idee e proposte, e ne abbiamo in cantiere molte altre; un *cantiere* che non si ferma e non si deve fermare mai perché *"nulla è permanente, tranne che il cambiamento"* (Eraclito, circa duemila anni fa). To be continued.

Vincenzo Giliberti  
Direttore Solution Design & Innovation  
Teleperformance Italy

Figure 3: Teleperformance Newsletter October 2014



On June 27<sup>th</sup> 2014 Teleperformance Italy has sent over two thousand employees and customer a newsletter on SENSEI Project, with reference to internal meeting that TP organized in Taranto on 25th and 26th May 2015.



Figure 4: Teleperformance Newsletter June 2015



### 3.2.4 Contacts and Cooperation

Table 3 shows the cooperation activities and external contacts taken during Period 2 of SENSEI project. Hereafter, the cooperation with Excitement project is described more in detail, including also the results of the experiments carried on possible application of the open platform to SENSEI.

Table 3: external contacts and cooperation activities Period 2

Type of activity	Main Leader (Partner)	Date	Place	Type of Audience: Scientific community	Size of Audience	Purpose / Justification / Outcomes	Relevance for SENSEI	Name of the attendees/speaker
Collaboration contacts	USFD	ongoing	Sheffield, UK	Scientific community		Regarding on specific SENSEI related tasks we have been talking to people at the University of Edinburgh and Indian Institute of Science, Bangalore, India.	WP5: linking and summarizing news article comments	Ahmet Aker
Collaboration with EHU (University of the Basque Country)	UESSEX	ongoing	-	Faculty members and researchers	6	We are collaborating with Ander Soraluze, Olatz Arregi and Xabier Arregi from the University of the Basque Country (EHU) on adapting our Coreference Resolution system, BART, to Euskara language (Basque).	Relevance for SENSEI WP4 on coreference domain and language adaptation, strengthening and complementing work on coreference for English, Italian and French.	Mijail Kabadjov, Udo Kruschwitz, Massimo Poesio, Ander Soraluze, Olatz Arregi and Xabier Arregi
Collaboration with Excitement team	UNITN	13/01/2015	Trento	SENSEI and Excitement from UNITN and FBK	6	Discussion on Excitement open platform and possible applications in SENSEI	Recommendation n. 6	Fabio Celli, Carmelo Ferrante, Evgeny Stepanov (UNITN), Roberto Zanolini, Simone Magnolini, Viviana Anasasi (FBK)
Visit to EJCM (Ecole de Journalisme et de Communication d'Aix-les-Bains)	AMU	30/01/2015	Marseille	Journalism students	10	Visit to EJCM: NLP class + presentation of the SENSEI to journalism students, discussion	Interest in SENSEI technology	Benoit Favre, Frederic Bechet

Type of activity	Main Leader (Partner)	Date	Place	Type of Audience: Scientific community	Size of Audience	Purpose / Justification / Outcomes	Relevance for SENSEI	Name of the attendees/speaker
Marseille)								
Meeting with EJCM students	AMU	25/03/2015	Marseille	Journalism students		Meeting with EJCM students to run a pre-pilot of the social-media use case to validate the approach	Pre-pilot of the social media use case to validate the approach	
Visit to Guardian	USFD	22/5/2015	London	Media Professionals: managing editor of Guardian, head of the Guardian comments section	5	Visit to Guardian. Presentation of WP1 and WP5 outputs to end users, Focus group discussion with end users	Relevance for SENSEI WP1/WP5; Feedback: Satisfaction with SENSEI approach to eliciting user requirements; Particular interest in SENSEI technology for clustering readers' comments in topic clusters; Agreed that a close research partnership could be mutually fruitful	Rob Gaizauskas, Mark Hepple, Jonathan Foster, Emma Barker





#### 3.2.4.1 Collaboration with Excitement

The UNITN team met the Excitement FBK team in Trento in January 2015. The goal of the meetings was to understand the technology and current state of development. UNITN discussed and shared the output of these meetings within the SENSEI consortium. Three internal activities spun off in order to get a feedback on the usability and performances of Excitement tools at that time. Three partners volunteered to evaluate the tools: UNITN, AMU and USFD. Below we collectively report on the outcome of these exploratory evaluation activities by the partners.

The Textual Entailment (TE) platform developed by the Excitement project may be used for a variety of NLP tasks such as Question Answering, Information Retrieval, Summarization. The platform is built on top of Apache UIMA Framework and it supports English, Italian, German. The performances on the TE tasks vary from data sets and languages. For instance, they have declared that on the RTE3 data set ( Italian ) the performance range between 60-70% and there is an additional 10% drop on speech transcriptions.

A possible application of the TE platform was identified in the speech use case of SENSEI for:

- Removal of duplicate sentences/utterances pre and post *summarization* (synopsis)
- Detection of sentences/utterances entailing the summary (resource evaluation)

however the current performance of the EXCITEMENT models are very low on out-of-domain data.

Another possible application the SENSEI team have investigated is for the social media use case, namely the sentence linking task. Given a newspaper article and a set of comments associated with it, relationships at sentence level need to be extracted as a preprocessing step for summary generation. For this, sentences need to be linked and then relations named. This relation between textual fragments can be seen as a textual entailment problem. The entailing and entailed text in TE are termed as text (T) and hypothesis (H), in other words, a human reading T would infer H to be true.

To link article-comment and comment-comment sentences, we assume comment to be **T** and article/comment to be linked to has the **H**. For instance

**Comment (T):** Not just cabinet members kept in the dark, but members of the national security council !

**Article (H):** Cabinet was told nothing about GCHQ spying programmes

To test the hypothesis, we evaluated the Excitement Open Platform (EOP) on the MultiLing-OpenForumS shared-task challenge designed by SENSEI. The OpenForumS data contains sentences linked between articles and comments and between comment and comment extracted from Guardian Newspaper, verified through crowd sourcing. From a set of 8 articles, 336 random valid links were selected for the experiments. (**T,H**) Different TE algorithms in the EOP platform were used for experiments on the RT3 dataset. The best F-score performance achieved for this task was 9.09 ( edit-distance method ) and a poor recall. Out of the 336 links (a small investigative experimental setup), only 12 were found by the system. The best performing model is based on Edit distance [3]. Last but not least the assumption that TE can be used to extract links may be wrong.



Altogether, although the TE is a promising approach, we did not find it effective for the SENSEI tasks.

### 3.3 Dissemination plan for Period 3

In this section we report on the dissemination plans for Period 3 which will reflect an increased activity due to the generation of the first results of technology development and evaluation.

UNITN plans to coordinate a journal paper on the SENSEI projects vision and intermediate results.

UNITN has a paper accepted and will participate to CLICit conference in December 2015.

UNITN, AMU, USFD and ESSEX will submit papers at peer-reviewed international conferences, such as ICASSP, LREC, INTERSPEECH, ACL, EMNLP. AMU will also continue its interaction with EJCAM, a journalism school, with the objective of teaching journalism students how they can benefit from natural language processing tools and about the SENSEI project in general.

USFD plans to have a paper in a peer-reviewed, international conference. In addition, an article presenting SENSEI results to general public will be written and published in an appropriate venue.

UESSEX will present a paper reporting on the shared task OnForumS to the Forum for Information Retrieval Evaluation 2015 (FIRE'15). We plan to submit a paper to a suitable outlet next year reporting on the work on domain adaptation for coreference. We also plan to submit a paper jointly with AMU to LREC 2016 reporting on the work on extending the Coreference Resolution system, BART, to French and likewise, a paper jointly with EHU on adapting BART to Basque. Also, a third paper for LREC presenting the datasets annotated for co-reference as part of SENSEI is within the plans.

Websays future dissemination plans include having a booth in the Smart City Expo World Congress 2015. Speakers and thousands of visitors will be together on one of the largest events about smart cities. Public opinion analytics is becoming a key issue in city participative processes and therefore SENSEI's work is very relevant to this community. The even in 2014 already showed clear interest from the community in SENSEI's potential and we expect interest will be higher this year.

TP has planned for Period 3 an increase in dissemination activities and in particular:

- a) **One newsletter** will be sent to employees (over 149.000 in 62 Countries) and customers of TP Worldwide;
- b) **Two newsletter** will be sent to employees (over 2.000) and customers of TP Italy;
- c) **Three articles** on the local, national or specialized press;
- d) **One video** that will be uploaded on YouTube and all blogs of TP with the answers given by WP leaders the following questions:
  1. Why do you think that Sensei objectives are important?
  2. What are most important technical challenges?
  3. What is the expected impact of Sensei?



4. How Sensei's technology and expected results could impact products and services in the future?

## 4. Exploitation activities

In this section we present the exploitation activities in Period 2 and we follow-up on the reviewers' recommendation ( n.9 in the list of recommendations).

### 4.1 Exploitation activities Period 2

In Period 2 our industrial partners (TP and Websays) have followed the plans presented in D7.2. In particular there have been two kinds of exploitation of SENSEI activities and outputs. The first kind is to exploit internally by improving the organization processes, products and services. The second kind is to attract and buy-in current and new customers into the benefit of SENSEI's vision and technology. In the following we report on the current year (Period 2) from both Websays and TP side.

#### 4.1.1 Internal Exploitation

Teleperformance has lead the development of the Agent Conversation Observation Form (ACOF) tool which provides Quality Assurance supervisors of contact center companies with a user friendly web interface to fill for each conversation:

- the monitoring forms "AOF - Agent Observation Form";
- the synopsis called "COF - Conversation Oriented Form".

The prototype of SENSEI ACOF provides automation in both forms: AOF and COF in further significant increases of performance and productivity. In a contact center as TP people dedicated to the QA team hear about 5% of total calls. This means that currently the remaining 95 % with all of its potential value in terms of knowledge on the behavior of the agents and the contents of the conversations is not even examined for obvious reasons of costs. Through the introduction of SENSEI ACOF technology in the contact center, it will be possible to greatly increase the coverage of the analysis, with no personnel cost increase. Consequently it is possible to improve the productivity and accuracy of quality assurance professionals and consequently boost the contact center performance. The synopsis generated by the systems SENSEI ACOF, coupled with the powerful full-text search systems like Elastic Search and Reporting ( Kibana ) would allow our customers for making direct business Intelligence on the content of the conversations.

SENSEI partners feedback is being used to improve WEBSAYS products and services. Some examples of technologies, which have already been taken into production, are:

- The asynchronous crawler developed specially for data collection presented in D2.2 and already presented in D7.2 has been constantly maintained and adapted as it has proven to be an important improvement for the WEBSAYS product. The asynchronous crawler reduces the time and cost of crawling URLs mentioned in posts, search results, etc.



- The parsers developed specially for SENSEI for blogs, forums and newspapers have been very productive when such data sources need to be acquired and have been constantly maintained and updated since they were deployed for D2.2.
- Data collection split (sharding), presented in D2.3, has represented a big performance improvement for large accounts like SENSEI. When the data collection is too large to be stored in a single machine Solr shards technology has been used to split the data collection into different machines. All the tools developed to this aim are already improving several WEBSAYS accounts with large data collections.
- Sorting data by Conversation Size, explained in D2.3, is helping WEBSAYS dashboard users to easily get posts with highest number of comments in a selected period of time. Several WEBSAYS clients have provided feedback confirming that this feature is very useful.

Since the beginning of the SENSEI project, a number of developments have led directly to improvements in Websays' pipeline and product offering.

In the first year of the project, the development of the asynchronous crawler for SENSEI (see D2.1) led to the widespread use of this crawler in Websays. Since its introduction in the main Websays pipeline in 2014, **all clients** benefit now from this development.

This is a back-end improvement, which enables new forms of crawling and indexing. This technology allows us to improve the data quality and freshness in two very important ways:

- Real time indexing of external links.
- Pagination of Comment pages.

In Period 2 of the project, the development of the "conversation size" metric for SENSEI (D2.2) opens a new venue in ranking content in Websays. Again, this was introduced in the main Websays pipeline, and today 50% of clients benefit from this development. Due to performance constrains, we cannot yet bring this feature to our largest costumers (in terms of volume of data), so R&D continues on this front.

This is a feature improvement which finds its way in several Websays tools:

- New "Sort By" option sorting by conversation size allows client to quickly view the social media items of most current impact and engagement.
- New "Top-10" report shows to our clients the top-10 items (sorted by conversation size) of each of their social media channels
- New Alerts: when then conversation-size metric grows beyond a certain threshold (increasing over time) the client receives an SMS and/or email alert.

#### **4.1.2 External Exploitation Events**

In this section we present the industrial events in which SENSEI project partners participated.

*CXV The Customer eXperience Value , May 27<sup>th</sup> 2015*



TP organized a large event in Milan (Italy), called “CXV The Customer eXperience Value - how to turn a negative experience into an opportunity to create value” and focused on multi-channel customer experience value. In this event TP presented to customers its products as well as Sensei Project and its opportunities. In this event the TP C-levels and the Professor Giuseppe Riccardi showed the Sensei’s goals to explore the real interest of Customers in these items. During this event has emerged strongly from many Companies present the need for multi-channel solutions (speech and web/social media) and analytics as the objectives of Sensei.

The main speakers at the event were:

- Paolo Righetti, Founder and CEO of GN Research, member of Global Management Team of Teleperformance WW, responsible of TP Added Value Analytics.
- Giuseppe Riccardi, University of Trento, Sensei Project Leader
- Gabriele Piva, CEO of Teleperformance Italy
- Gabriele Alban, Vice President Account Management, IT & Security of TP Italy
- Mauro Fanfoni, Senior Vice President Marketing and Innovation Europe Eni
- Anna Imperadori, Vice President Operations Barclays
- Marco Corradino, Founder Volagratis
- Biagio Stasi, Senior Digital and Strategy Director Hearst Magazines

The chairman of the event was:

- Marco Montemagno, Digital specialist and founder of Blogosfere.

The participants at the event were C-Levels, Sales Managers, Marketing Directors, Customer Care Directors and Social Managers of Italian and European Companies in the following industries:

- Telco, High Tech, IT: 4
- Travel, Transportation, Automotive, Logistics: 4
- Financial Services, Bank, Insurance: 8
- Energy & Utilities: 6
- Media & Entertainment: 4
- Retail, E-Retail: 4
- Other: 10





Figure 5: Marco Montemagno interviews Giuseppe Riccardi during “CXV The Customer eXperience Value”



Figure 6: Gabriel Piva, CEO of TP Italy, Gabriele Albani, VP of TP Italy and Paolo Righetti, Founder and CEO of GN Research during “CXV The Customer eXperience Value”

### *The Smart City Expo , November 18<sup>th</sup> 2014*

The Smart City Expo World Congress<sup>i</sup> took place in Barcelona in November 2014. It brought together over 400 cities around the world, 242 companies, 400 speakers and more than 10.000 visitors. The Smart City Expo is definitely the most important global event on the smart cities' calendar.



Figure 7: Smart City Expo World Congress



The event was addressed to municipalities, enterprises, entrepreneurs, research centres, universities and other public or non-governmental organizations or consortia (public-private) to spread ideas, research, insights and innovative solutions around the Smart City concept.

The congress program was composed of 5 keynotes, 6 plenary sessions and 42 breakout sessions. It was the meeting point for cities around the world to share their visions and solutions towards becoming more sustainable and easy to live in.

Websays participated in the congress with a stand in which multiple presentations and demos were carried out. Websays showcased the SENSEI project goals and early prototypes on a large screen and brochures. Many relevant people visited the booth and heard about SENSEI technology including:

- European projects: FIWARE, FUPOL
- Corporations: Oracle, Microsoft, ScytI, Telefónica
- Institutions: Barcelona City Council, (see below for more)

The following pictures show the Websays conference booth and a demo of the Social Media document analysis ( shown at 1<sup>st</sup> Period Review meeting in Luxembourg) , presented to an executive of ScytI, leader in online and electronic voting, interested in extending their services to analytics about opinion during an event.



Figure 8: SENSEI at Smart City Expo World Congress

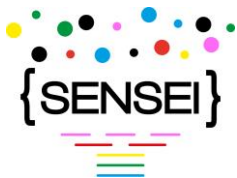


Figure 9: SENSEI Demo presented to ScytI

Following the SmartCity Congress Websays had meetings with several city councils and agencies pitching the Websays City Intelligence Solution. In each of these meetings SENSEI was presented and early prototypes shown demonstrating the importance of conversation analytics for their needs according to the dissemination task presented in DoW<sup>2</sup>. The most important of these meetings are: London Borough of Hackney (GB), Brighton (GB), Dublin (IE), Barcelona (ES), Tarragona (ES) and Sant Feliu (ES).

Despite the early stage of SENSEI's prototypes, they were very useful to position Websays as a world-leading company in opinion and conversation analytics with long-term R&D tackling our client's hardest issues.

*Listening to Social Networks<sup>3</sup>, March 24, 2015*

It was a workshop by Catalonia Local Administration Consortium (LOCALRET). Websays presented the present and future analytics technology for understanding citizens and carry out actions based on this analytics. SENSEI was also presented and a prototype shown.

## 4.2 Exploitation Plan

Moving forward the industrial partners have initiated the elaboration of an exploitation plan that identifies products and services that can be presented to their organizations to improve the portfolio offering or improving the internal costs of process development and execution.

<sup>2</sup> Industry dissemination and exploitation, DoW part B, Section 3 Impact

<sup>3</sup> <http://www.localret.cat/agenda/taller-localret-us-i-escolta-de-les-xarxes-socials/>





### 4.2.1 Conversation Analytics

The actors that will get most benefits from Sensei technologies include, but are not limited to:

- the **human resource department** of contact center will be able to deliver more focused training;
- the **Operation department** of contact center will rely on more skilled and motivated agents;
- the **Marketing, Sales and Account Managers** can share with our Clients the business intelligence results;
- the user-end **Clients** who will *hear* the customers' voice in the form of readable and useful summaries.



In addition, as the Sensei technology become mature and near real-time, it will monitor the Agents during the call and provide suggestions **directly to the Agents** to improve the conversation.

The target clients of Sensei technologies are both **small and large companies**; the first can count on Sensei monitoring without additional QA personnel costs while the seconds can take advantages of the benefits according to their budget.


#### 4.2.1.1 The assessment of the state of the art: Market trends and relationship with SENSEI

To assess the state of the art in the contact center innovative technology solutions, TP is continuously monitoring the market trends and in particular analyze the results of the periodicals different reports issued by the most leading research and advisory company.

Table 4: The reports on contact center technology trends, and the relationship with Sensei project analyzed by TP Italy

Company	Short Company Profile	Report examined
	<p>Gartner is the world's leading information technology research and advisory company.</p> <ul style="list-style-type: none"><li>• More than 960 expert analysts cover 1,200 topics across the IT landscape</li><li>• Gartner analysts are based in 26 countries and speak 47 languages</li><li>• Gartner analysts have an average of 12 years' experience in their specific field</li></ul>	«Magic Quadrant for Customer Management Contact Center BPO», 2014, Gartner, TJ Singh e Brian Manusama
	<p>Ovum is part of Informa Group, the world's leading provider of business intelligence.</p> <ul style="list-style-type: none"><li>• 220+ country market forecasts and KPI's</li><li>• 10,000,000+ market data points</li><li>• 180 analysts, each with on average</li></ul>	“CRM Outsourcing Business Trends 2014”, OVUM, Peter Ryan.



	more than 10 years industry experience.	
	Cerved is the Italian's leading provider of business intelligence	"Smart Call Center", 2014, Databank Gruppo Cerved, Elena Pavan

Following Table 5 shows the main recommendations by Ovum Consulting report 2014 and relationship with SENSEI.

Table 5: The recommendations by Ovum Consulting report and relationship with Sensei


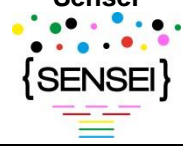
Recommendations by Ovum Consulting report 2014 	Sensei 
<ul style="list-style-type: none"> <li>▪ <b>"Emphasize channel capabilities"</b> <i>This year's survey shows that enterprise contact centers are deploying a very wide range of contact mechanisms for their end users. In order to move business to third-party delivery and compete with other outsourcers, vendors are recommended to ensure that their channel offerings reflect the needs of consumers related to targeted country and vertical markets, in order to ensure maximum relevance to prospects."</i></li> </ul>	<p>The project <b>Sensei is natively multi-channel with an emphasis on voice channels and social media.</b></p> <p>The goal of Sensei is to provide a unified data view of "conversations", both from speech dialogues and online (social media) dialogues.</p>
<ul style="list-style-type: none"> <li>▪ <b>"Customer analytics is a must"</b> <i>Outsourcers must do their best to deepen their value-added offerings in order to avoid margin erosion. It was clear in this year's survey that in-house contact center headaches relate to being able to get the broadest view of the enterprise end user, so it is recommended that outsourcers use this to their advantage by offering full analysis packages in which data is extracted, analyzed, and interpreted. This will not only add value to the outsourcer's bottom line in terms of higher-margin work, it will also deepen relationships with clients."</i></li> </ul>	<p>The other main objective of Sensei is to <b>provide an efficient tool for analyzing the conversations between customers and agents and agents' behavior to improve the perceived quality and the performance of the contact center .</b></p>

Table 6 shows the contact center outsourcing trends by Ovum Consulting from the report 2014.

Table 6: The contact center outsourcing trends by Ovum Consulting report

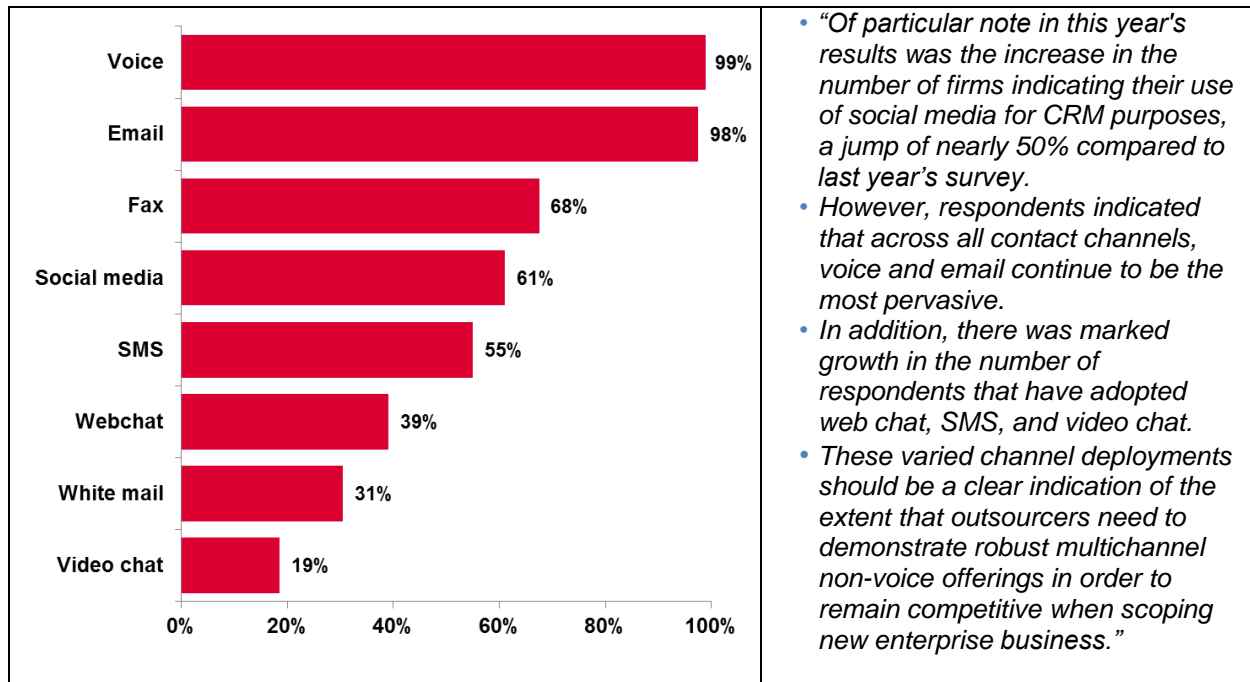


Table 7 shows the key factors identified by Gartner and relationship with SENSEI project.

Table 7: The contact center outsourcing trends by Ovum Consulting report 2014.

Contact center market: the key factors by Gartner, 2014	
<p>The <b>selection of competitors</b> will emerge the most competitive and <b>able to offer services with high added value</b> (amplitude of services and related activities).</p> <ul style="list-style-type: none"> <li>• <b>Managing multi-channel</b> push to use more and more massive of the new instruments of contact: the appropriate technology allows us to offer a more personalized service and to satisfy the user.</li> <li>• <b>Progressive digitization of the market</b> and targets set up by Digital Agenda and the possible development of new services interaction with the PA</li> <li>• <b>Automation technologies and speech recognition</b>, automatic systems response to no human activities and services less people intensive</li> <li>• Opportunities arising from new technologies and <b>integrations with social network</b> world and mobile app and web communication tools.</li> </ul>	<p>As mentioned above the project <b>Sensei is natively multi-channel with an emphasis on voice channels and social media.</b></p> <p>The objective of Sensei is to provide <b>an efficient tool with summary automation technologies (SENSEI ACOF) and integrate the speech technologies with the social media technologies and show the creation of value.</b></p>

These recent indicators of authoritative sources are extremely important and they confirm the market opportunities of the SENSEI project, whose objectives are among others:



- increase the competitiveness of companies, enabling them to offer services with high added value and managing multi-channel;
- increase the use of automation technologies and speech recognition, automatic systems response to no human activities;
- increase the integration with social network.

#### 4.2.1.3 The exploitation roadmap for the Contact centre business

In the first six months of Period 3 TP will realize a Survey, as described in the table below, to present Sensei technologies and potential benefits to find out what Clients think and try to find a partner interested in a Pilot.

Table 8: The Survey action plan

ID	Activity	Description
1	Survey Definition	TP will define: <ul style="list-style-type: none"> <li>• the survey contents;</li> <li>• the survey target;</li> <li>• the survey timing;</li> <li>• the survey channels: face to face, meeting, email, web, phone call with agents, automatic phone call, IVR;</li> <li>• the output reports;</li> <li>• the test plan.</li> </ul>
2	Survey Set Up	TP will prepare: <ul style="list-style-type: none"> <li>• the detailed list of questions and possible answers;</li> <li>• the list of recipients;</li> <li>• the activities planning;</li> <li>• the technology implementation of the selected channels;</li> <li>• the report templates;</li> <li>• the acceptance full test.</li> </ul>
3	Survey Deployment	TP will deploy the Survey and begin to collect data.
4	Survey Reporting	TP will release the reports of Survey.
5	Survey Analysis	TP will analyze the results of the Survey.

In the Period 3 TP will organize an event directly to reach potential customers/user of Sensei outputs, in one of its European subsidiaries (twenty European countries). The event will be attended by numerous C-Level, Sales Managers, Marketing Directors, Customer Care Directors and Social Managers of European Companies but also but also with Public Administration / Government.

Table 9: The Event action plan

ID	Activity	Description
1	Event Definition	TP will define: <ul style="list-style-type: none"> <li>• the Event contents;</li> <li>• the Event target;</li> <li>• the Event timing;</li> <li>• the Event location;</li> <li>• the Event speakers;</li> <li>• the channels to promote the Event : email, web, phone call.</li> </ul>
2	Event Set Up	TP will prepare: <ul style="list-style-type: none"> <li>• the detailed list of contents and the agenda;</li> <li>• the list of attendees;</li> <li>• the activities planning;</li> </ul>



		<ul style="list-style-type: none"><li>the channels to promote the Event;</li></ul>
3	Event Deployment	TP will host the Event and collect feedback.
4	Event Reporting	TP will release the report of Event.
5	Event Analysis	TP will analyze the results of the Event.

The goal of the survey will be to identify and match clients' needs and expectation with SENSEI outputs and technologies. Last but not least the survey will allow TP to follow-up with clients that are forward-looking and ready to follow SENSEI's innovative technology.

## **4.2.2 From Social Media to Newspaper companies**

### *4.2.2.1 Direct Exploitation of SENSEI R&D into Websays Products*

Since the beginning of the SENSEI project, a number of developments have led directly to improvements in Websays' pipeline and product offering.

In the first year of the project, the development of the asynchronous crawler for SENSEI (see D2.1) led to the widespread use of this crawler in Websays. Since its introduction in the main Websays pipeline in 2014, **all clients** benefit now from this development.

This is a back-end improvement, which enables new forms of crawling and indexing. Because of its nature, the client is not perceive this feature directly. However, this technology allows us to improve the data quality and freshness in two very important ways:

- Real time indexing of external links.
- Pagination of Comment pages.

In Period 2 of the project, the development of the "conversation size" metric for SENSEI (D2.2) opens a new venue in ranking content in Websays. Again, this was introduced in the main Websays pipeline, and today 50% of clients benefit from this development. Due to performance constrains, we cannot yet bring this feature to our largest costumers (in terms of volume of data), so R&D continues on this front.

This is a feature improvement which finds its way in several Websays tools:

- New "Sort By" option sorting by conversation size allows client to quickly view the social media items of most current impact and engagement.
- New "Top-10" report shows to our clients the top-10 items (sorted by conversation size) of each of their social media channels
- New Alerts: when then conversation-size metric grows beyond a certain threshold (increasing over time) the client receives an SMS and/or email alert.

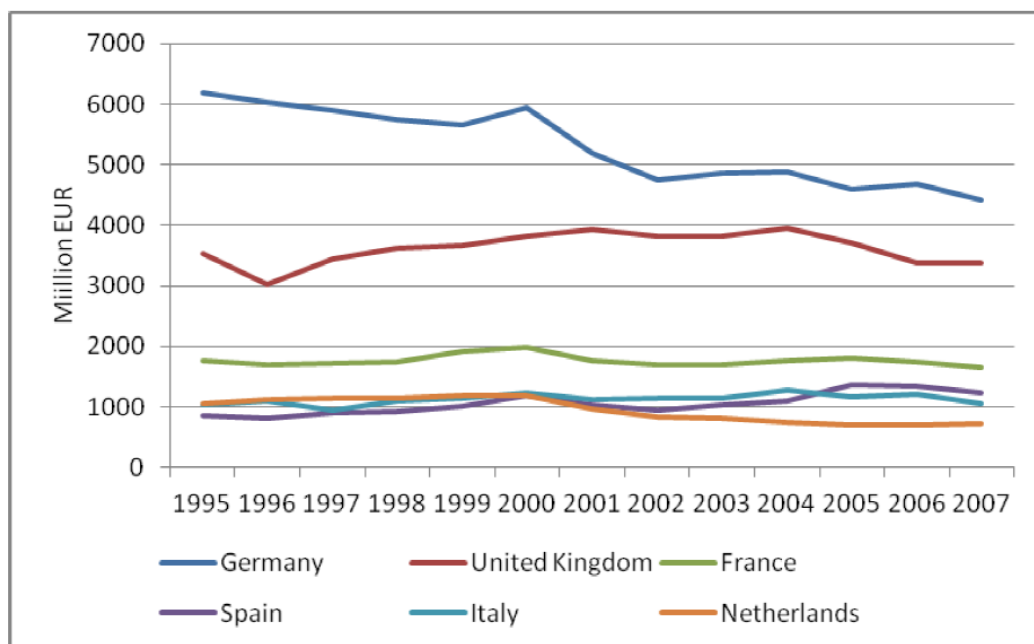
### *4.2.2.2 Market trends and relationship with SENSEI's results*

The Newspaper Market Industry has been tremendously slow at realizing the impact of the Internet in people's needs for information. There have been little changes in the industry over the last 15 years, as the information landscape changed radically changed under their feet.

Up until today, global newspaper revenue depends for more than half (57%) on advertising and for 43% on direct income from readers who have a subscription or buy single copies (PwC, 2009;

OECD, 2010). Online advertising revenue is only slowly increasing and worldwide accounted in 2009 for approximately 4 of newspaper publishers' revenues and 6% of total advertising revenue (OECD, 2010). Both of these sources of income are under serious threat.

European Publishing companies generate an overall revenue of approximately 50 bln€. However their annual average decline in the average firm size (-15.9%) and in the number of employees (-1.4%) clearly shows their current state of crisis.



Source: TNO, 2011, based on Eurostat.

Figure 10: Six largest EU Member States in terms of share in total EU value added in newspaper publishing, period 1995-2007

Figure 14 shows how the total valued added of the six EU Member States with the largest share in value added in the publishing of newspapers developed between 1995 and 2007. After 2000, the value added of the German and Dutch newspaper publishing industries decreased with respectively 1.5 billion and 467 million EUR to a total value added of 4.4 billion and 720 million EUR respectively in 2007.

In the recent years the decay seems only to have accelerated (although we do not have figures to back this) with the rapid growth of Internet usage hurting the newspaper industry in several fronts: fragmenting young audiences and brining new online players such as Google or Yahoo News.

In the past 6 years the most forward-thinking newspapers have been experimenting with novel ways to “bring people in”, to engage with their audience in new ways not available before the Internet. The solutions have been:

- Private Blogs: spaces where journalists can extend their news and give a more personal view to better engage with their audience.
- Infographics and Interactive Visualizations: to better explain the data.





- User Generated Content: allowing readers to provide pictures, videos or even stories to the newspaper.
- Article Commentaries: spaces where the audience can directly comment on the articles
- Social-media driven news and editorials: listening to the social media trends and events similarly to how they would listen to the “streets” events.

All of these attempts have been partial, with both positive and negative effects on the audience and the advertising revenue.

SENSEI's social media use cases have taken the last two points: commentaries and social-media driven news. In consultation with The Guardian, a number of prototypes were developed and evaluated.

Although the results from these interactions are preliminary, there seems to be a clear need for better technology that allows journalists to

- Better browse and summarize article commentaries, their main trends, controversies, etc.
- Better browse and summarize social media (in particular Twitter and Facebook), their main trends, controversies, etc.

In the following sections we provide a first exercise in developing an exploitation plan for these technologies.

#### *4.2.2.3 Technology products for the Exploitation Plan*

It is clear from the first evaluation of SENSEI NewsPaper prototypes that there is a clear interest by the Newspaper industry in advanced technology for summarization and visualization of user comments and ongoing social trends. We develop here a simple business model which will serve as a basis of discussion with The Guardian and other newspapers and the prototypes mature and their added value to the companies becomes clearer. The revenue model would be cloud SaaS, with a subscription fee structured around the following parameters: volume (in terms of posts) and user licenses. The actual pricing will depend on the final products developed after the prototype phase, and in particular will depend on the added value that these products bring to journalists.

Although it is hard at this early stage to estimate the revenue per contract, based on current experiences with Marketing Agencies and Newspapers we estimate a monthly recurrent revenue (MRR) for a single NewsPaper of 50k-100k € per year, and for a publishing group of 100k to 250k€ per year. Assuming they keep this service an average of 3 years (being conservative), this leads to a lifetime value (LTV) of 150k-750k€, although it is hard at this point to determine if the average LTV will be closer to the lower or the higher end of this range. Given the very high gross margin for this type of SaaS business (around 80%), this leads to a very profitable business.

The cost of customer acquisition (CAC) is likely to be high; requiring extended trial periods and negotiations with the publishing groups. This is a risk to take into account and a factor that implies high Marketing costs and slow initial growth. Nevertheless, even with a relatively long customer acquisition time (e.g. 8 months), the LTV:CAC ratio would be high, well above the typical recommended threshold of 3.



The Road to Market can be structured in two phases: an initial market test where the road will be based on direct connections to European Newspapers, and an expansion phase where the road to market will be through traditional SaaS methods: targeted email and web campaigns and specialist shows and events. In this second phase there is no reason to restrict ourselves to the European market, since all Newspapers around the world have the same needs. The only constrain in terms of market expansion is the languages supported by the technology.

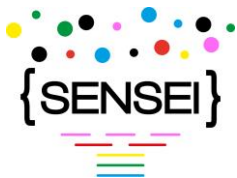
#### Development Plan

The cost and time needed for moving from prototypes to a final first version ready for the market are estimated (grossly) as follows:

- Front-End: 50K€
- Back-End: 100K€
- Design: 50K€

The required development time is estimated in 12-18 months. This shows that with a relatively small investment in capital and time we would be ready to launch a product world-wide. More importantly, we see that the required R&D investment in this product is marginal ( $< 1$  LTV). However the overall expected return of the investment (ROI) depends completely on the expected sales and the CAC. It is too early to develop realistic sales and CAC figures, until SENSEI research does not mature and the prototypes are not validated.





## 5. Managing Intellectual Property

The Consortium Agreement provides a legal framework for protecting and managing IPR. Following these guidelines, the Consortium is managing intellectual property issues that arise through the work of individual partners or the consortium as a whole.

With respect to data distribution of the data collections presented in D2.1, D2.2 and D2.3:

LUNA and DECODA speech collections were distributed under their pre-existing IP protection mode, which consisted in releasing a small sample and requesting signed forms for access and (restricted) use of the collection.

Social Media Collection v1.0 and v.2.0: following an evaluation of the available options and discussion, it was decided to release a small sample of the full content as well as full URL list of sources as a fallback and secure option (see IPR survey presented in D8.4).

Improvements to the Websays processing pipelines and storage system remain Websays IP, for the moment treated as trade secrets and not released to the public, although the general principles have been shared with SENSEI partners to make them benefit from Websays know-how. On the other hand, crawlers for the specific newspapers developed for SENSEI will be made public and free.

With respect to the Speech Scenario one of the evaluation tool, the ACOF (Agent Conversation Observation Forms), has been made public in a definitive version in deliverable D1.2.



## 6. Conclusions

Period 2 has been a crucial period in terms of bootstrapping the dissemination activities and reporting externally on the vision and intermediate results. We believe we have set the ground to leverage on these intermediate results and further improve SENSEI's visibility and impact in Period 3.



## References

- [1] Statistical, Ecosystems and Competitiveness Analysis of the Media and Content Industries: The Newspaper Publishing Industry. Andra Leurdijk, Mijke Slot, Ottilie Nieuwenhuis. JRC Report EUR 25277 EN. 2012.
  - [2] <http://multiling.iit.demokritos.gr/pages/view/1531/task-onforums-data-and-information>, accessed on June 2, 2015
  - [3] Ido Dagan, Oren Glickman and Bernardo Magnini. The PASCAL Recognising Textual Entailment Challenge, p. 2 in: Quiñonero-Candela, J.; Dagan, I.; Magnini, B.; d'Alché-Buc, F. (Eds.) Machine Learning Challenges. Lecture Notes in Computer Science , Vol. 3944, pp. 177-190, Springer, 2006
  - [4] <http://excitement-project.eu/>, accessed on June 2, 2015
-