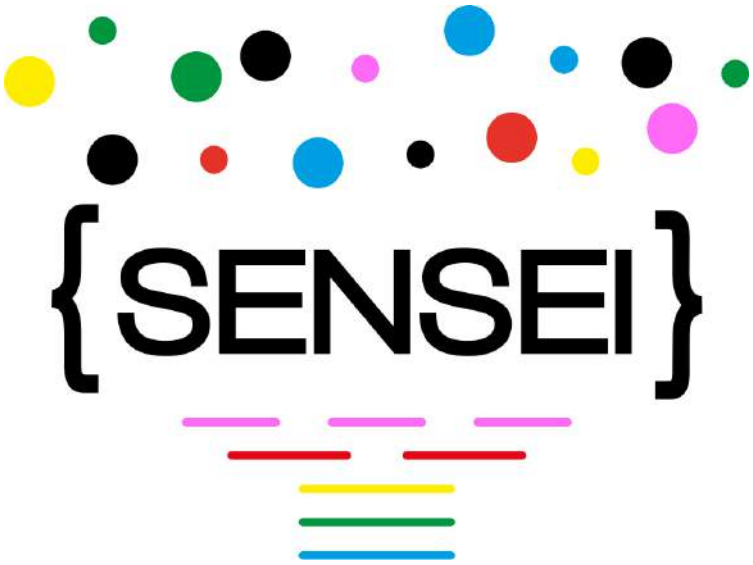

D2.1 – Intermediate Data Collection Samples

Document Number	D2.1
Document Title	Intermediate Data Collection Samples
Version	3.0
Status	Final
Work Package	WP2
Deliverable Type	Report
Contractual Date of Delivery	30.04.2014
Actual Date of Delivery	30.04.2014
Responsible Unit	WEBSAYS
Keyword List	Data Collection Samples, Schema
Dissemination level	PU

Editor

Hugo Zaragoza (WEBSAYS)

Contributors

Fabio Celli	(UNITN)
Giuseppe Riccardi	(UNITN)
Benoit Favre	(AMU)
Robert Gaizauskas	(USFD)
Mijail Kabadjov	(UESSEX)
Vincenzo Lanzolla	(TP)
Marçal Juan	(WEBSAYS)

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	1/2/2014	draft	Hugo Zaragoza	Table of Content
0.2	10/2/2014	Draft	Hugo Zaragoza (Websays)	Draft of each section added
0.3	10/03/2014	Draft	Marçal Juan (Websays)	Added work overview
0.4	12/03/2014	Draft	Vincenzo Lanzolla (TP)	TP data description and annexes
0.5	15/03/2014	Draft	Fabio Celli, Giuseppe Riccardi (UNITN)	Schema Model modifications
0.6	17/03/2014	Draft	Rob Gaizauskas (USFD)	Social Media & Schema Model modifications
0.7	26/03/2014	Draft	Mijail Kabadjov (UESSEX)	Suggestions
0.8	31/03/2014	Draft	Hugo Zaragoza (Websays)	Data model description, annex and corrections Modified as suggested by partners in discussion.
1.0	07/04/2014	Draft	Hugo Zaragoza (Websays)	Consolidated version for review process
1.1	07/04/2014	Draft	Benoit Favre (AMU)	Scientific Review. Comments and suggestions
1.2	07/04/2014	Draft	Elisa Chiarani (UNITN)	Quality check completed. Some remarks and use of the template
2.0	10/04/2014	Draft	Hugo Zaragoza (Websays)	Made necessary corrections and rewrites as per remarks. Updated and consolidated version after review
2.1	22/04/2014	Draft	Giuseppe Riccardi (UNITN)	Review. Suggestions and minor remarks
2.2	24/04/2014	Draft	Hugo Zaragoza (Websays)	Consolidated version
2.3	28/04/2014	Draft	Elisa Chiarani, Giuseppe Riccardi (UNITN)	Quality Check. Some minor Remarks
2.4	29/04/2014	Draft	Adam Funk (USFD)	Quality check from a native speaker of English. Some minor changes
3.0	29/04/2014	Final	Hugo Zaragoza (Websays)	Final version ready for submission

INDEX

Executive Summary

This deliverable provides human conversation data samples from speech and social media use cases to partners for validation and feedback in preparation for the final complete data collection to be delivered in deliverable D2.2.
This document provides information on the available corpora (speech use case) as well as data model and specifications of the data (social media use case) to be acquired, statistics and data sources and access.

1. Overview

The preliminary speech data collection is composed of three collections: DECODA and LUNA (pre-existing collections) and a Teleperformance collection from TP partner. The social media collection has been crawled from different Internet sources.
A number of intermediate tasks were necessary to achieve this deliverable, and are briefly described in this document:

- Study (with WP1 and WP6) select appropriate data sources;
- Extraction and documentation of Teleperformance data;
- Definition of a Social Data Schema;
- Adaptation of the Websays crawler to the required sources;
- Evaluation of the obtained data.

1.1 Approach

The ultimate goal of WP2 is to provide a unified data view of “conversations”, both from speech dialogues and online (typed) dialogues. This however requires a high level of abstraction from the raw data, which is not readily available; indeed, building such an abstraction is one of the main objectives of the SENSEI project.
After much discussion and experimentation we have decided to proceed as follows. First, D2.1 should provide views on the data in a way that the full original data could be reconstructed. Second, additional annotation on the data should be provided by other WPs in the form of stand-off annotations on these views. Third, a mapping between data schemas should be developed to achieve a unified conversation schema (partially for D2.2, complete for D2.3).
To this end, for deliverable D2.1 we have collected several collections from several different sources, documented their data schemas, and have begun to draw parallels between the different data sources.

1.2 Data Access

The initial data set contains three parts and over 400k items. Exemplars from the Social Media and LUNA collections are provided for public online access from the SENSEI web site, together with this document, which provides an overview of the data and instructions about how to request the entire data sets. The method of data acquisition and usage is discussed in D8.2 – Ethical Issues Plan. Here we recall some of the most relevant information fully contained in deliverable D8.2.

For the Social Media collection, the website provides a data bundle for D2.1: a small sample of 1000 social media items from the Social Media collection, together with the entire list of public URLs of the data crawled for this collection. The entire collection (as well as individual parts of the collection) can be made available to the public upon e-mail request to "sensei-data@list.disi.unitn.it" <sensei-data@list.disi.unitn.it>.

For LUNA the data we provide a small complete sample; the entire collection is distributed as-is to partners for evaluation and annotation through the data sharing agreement provided in the Ethical Issues Plan (D8.2).

For DECODA the data we provide a small complete sample. The entire collection is distributed by SLDR/Ortolang (http://crdo.up.univ-aix.fr, ID: http://sldr.org/sldr000847). Researchers or practitioners may get access to the annotated corpus of human conversations free of charge by accepting the SLDR/ORTOLANG license.

In the case of the Teleperformance the data is available to the partners internally already, but will not be ready to the public until D8.3 is finalized.

For partners, a SVN data repository has been setup on one of the SENSEI servers containing all the data for easy access.

The Websays Dashboard can also be used to browse the Social Media portion of the data:

```
https://websays.com
user: sensei_demo1
password: sensei_demo1
```

2. Speech

2.1 Decoda Collection

The French RATP-Decoda collection consists of call-center recordings from a public transportation agency (RATP), with several levels of annotation: word transcripts, part-of-speech tags, dependency parses, named entities.
It contains 1514 human-human dialogs between a customer and one or more agents. The dialogs were collected as part of the French Agence Nationale de la Recherche (ANR) Decoda project. They cover topics such as lost-and-found, fares information, complaints or itinerary, route queries, etc. The corpus is particularly coherent because it covers a full day of calls on all types of calls. All calls are anonymized.
The dialogs consist of 56,660 turns, 600,003 tokenized words, with a lexicon of 11,375 different tokens. The audio recordings cover about 100 hours of speech. Part-of-speech tags and dependency parses follow the P7 French Treebank guidelines adapted to speech. The annotation was bootstrapped with automatic systems, carefully manually corrected on a subset (the gold subset), and corrected through rule propagation on another subset (the silver subset). Transcriptions, disfluencies and named entities are annotated according to the Ester guidelines, adapted to the public transportation domain. Dialog-level semantic categories are distributed as follows:

Table 1: Distribution of Dialog-level semantic categories

Call Type	%
Traffic information	22.5
Route planning	17.2
Lost and Found	15.9
Registration card	11.4
Timetable	3.6
Ticket	4.5
Specialized calls	4.5
Empty	3.6
New registration	3.4
Price info	3.0

In addition, Automatic Speech Recognition (ASR) transcript from a baseline ASR system is provided for all the dialogs.
All speech transcripts input to SENSEI are in transcriber trs (xml) format. Annotations are in a table format similar to that of the CoNLL'07 shared task format.

2.2 Luna Collection

The Italian LUNA Corpus is a collection of 572 human-human dialogs in the hardware/software help desk domain. The dialogs are conversations of the users and operators involved in

problem solving. The dialogs are organized in transcriptions and annotations defined within FP6 LUNA Project. The dialogs were annotated at different levels: words, turns, attribute-value pairs, predicate argument structure and dialog acts.

The annotation at word level consists of lemmas, part-of-speech tags and morpho-syntactic information following EAGLES corpora annotation. Attribute-value annotation makes use of predefined ontology of domain concepts and their relations. Predicate argument annotation is based on FrameNet model. Dialog act annotation was inspired by DAMSL, TRAINS and DIT++ and is used to mark intentions in an utterance. Discourse relation annotation was done following Penn Discourse Treebank (PDTB) approach.

The general process of annotation can be seen in the figure below. Dialog act and attribute-value annotation is done on segmented dialogs at utterance level. However, predicate argument annotation requires POS-tagging and syntactic parsing. This was done semi-automatically using the Bikel parser trained on an Italian corpus with subsequent manual correction.

Table 2 below provides general statistics on the LUNA Corpus, such as number of dialogs annotated at each level, as well as token and turn counts.

Table 2: general statistics on the LUNA Corpus

	Size
Transcribed and annotated at AV level	572 dialogs
Total time in min.	1,790 min
Total number of tokens	207,200
Total number of turns	26,638
Total number of chunks	156,064
Total number of concepts	46,027
Total number of different words	9,532
Annotated at DA level	81 dialogs
No. of dialog acts annotated	3,203
Annotated at PS level	78 dialogs
No. of frames annotated	4,367
No. of frame elements annotated	4,777
Discourse Relation Annotation	60 dialogs
No. of Relations	1,606

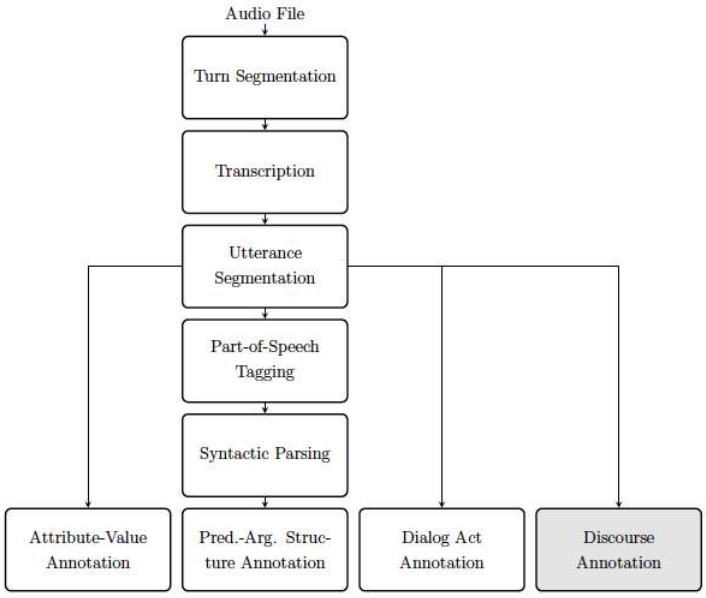

Figure 1: LUNA annotation process

Additionally ***LUNA Human-Human Corpus*** has gone through additional quality control/cleaning procedure.

1. Attribute-value annotation (concept ontology) is normalized;
2. Due to the nature of the hardware/software Help Desk domain, the corpus contains words borrowed from English. Thus, the corpus is corrected for misspelled words via semi-automatic procedure (automatic detection of misspelled words with human decision on the correct form);
3. Since LUNA Corpus contains sensitive private information, such as personal names, phone numbers, etc., which is protected by Italian privacy laws; the corpus is anonymized. A special attention is given to preserve the distribution of token within anonymized concept values. However, transcriptions and predicate-argument structure layer is not anonymized due to different segmentation and tokenization, which makes it difficult to distinguish sensitive vs. non-sensitive data (see the Figure above on annotation process).

2.3 Teleperformance Data

The Call Quality Monitoring tool is a Teleperformance web tool where Quality Assurance professionals (listeners) evaluate the monitored call based on a set of items and sub-items defined by a Quality Manager. Some items and sub-items are standard and are applied for each client and their so-called “campaigns”. Other items and sub-items are defined following campaign-specific and client-specific requirements.
Below is a simplified data model.

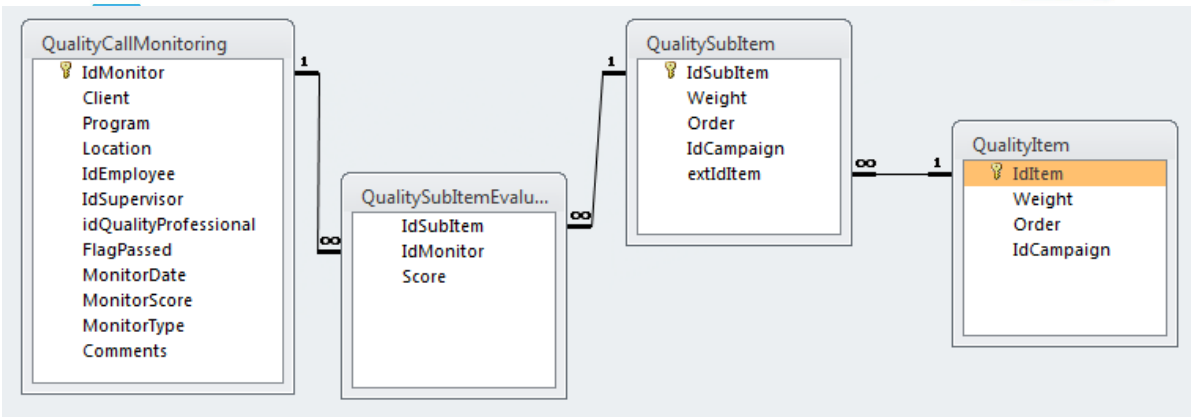

Figure 2: Call Quality Monitoring

QualityCallMonitoring: contains the overall evaluation of the Call Quality Control.

Table 3: Call Quality Monitoring

Field	Description
IdMonitor	Unique Identifier of the Call Quality Monitor
Client	Name of the Client
Program	Name of the Campaign/Project
Location	Teleperformance's site where the campaign is managed
IdEmployee	Unique identifier of the Agent being listened
IdSupervisor	Unique identifier of the Manager's Agent being listened
idQualityProfessional	Unique identifier of the Quality Assurance Professional listening and evaluating the call
FlagPassed	Flag that indicates the outcome of the Monitor: 'Y' if the call passes the minimum quality requirement(according to TP standard and client needs) 'N' otherwise
MonitorDate	Date of the monitor
MonitorScore	Final Score calculated as the weighted sum of all subitems
MonitorType	Identify the type of campaign/project: 'IB' is the value for Inbound Campaign 'OB' is the value for Inbound Campaign
Comments	Free text field where the quality assurance professional write some comments

QualityItem: catalog of evaluation item (for the campaign).

Table 4: Quality Item

IdItem	Unique Identifier of the item
Weight	Weight of the item

Order	The items are displayed according to their order
IdCampaign	Unique identifier of the Campaign

QualitySubItem: catalog of evaluation sub-item (for the campaign).

Table 5: Quality Sub-Item

IdSubItem	Unique Identifier of the sub-item
Weight	Weight of the item
Order	The sub-item are displayed according to their order
IdCampaign	Unique identifier of the Campaign
extIdItem	External key to the related Item

QualitySubItemEvaluation: score of each sub-item.

Table 6: Quality Sub-Item Evaluation

IdSubItem	Unique Identifier of the sub-item
IdMonitor	Unique Identifier of the Call Quality Monitor
Score	Score of the sub-item

The standard items such as politeness, listening attitudes, empathy, track the agent opening and closing of the call, language usage, etc. In addition, the Call Quality Evaluation Form contains one or more free text fields where the Quality professionals write notes or summaries about the overall evaluation, strength and weakness points of the agents and any other information that may be useful.
An evaluation form example is provided in Annex II.

2.3.1 Preliminary Data Set

An inbound customer care campaign for an Italian company in the Energy industry was chosen as the source of the preliminary data set.
The preliminary data set for this deliverable consists of 1063 records, 7 evaluation items and 23 evaluation sub-items.

3. Social Media

3.1 Data Schema

Prior to data acquisition, we have studied and analysed the target data sources to determine the type of metadata and relationship data that was necessary to extract and store.

We decided to consider as the unit of publication a unit smaller than the page or the document, in order to deal with comments. We denote a “clipping” (or “mention”) the unit of publication and indexation. A mention represents for example a post (such as a blog post, a Facebook post, a Tweet or Retweet) or a comment associated with a post.

Mentions are represented in a way that the entire conversation can be reconstructed: they are indexed by author and the various available author_IDs (e.g. the Facebook apiAuthorID), date of publication, position in the comment thread, etc. They maintain pointers to their parent post (e.g. in the case of a comment). Furthermore they contain additional meta-data (e.g. the number of likes).

Two types of pointers are recorded: postID and parentID. The postID pointer allows for fast retrieval of elements within a post (without the need of recursive calls). The ordered-tree structure of comments to posts and comments to comments is preserved by the parentID and the position pointers. This is illustrated in Figure 3. The Complete Data Schema is provided in Annex I. Annex I also provides an alignment between speech and social media schemas.

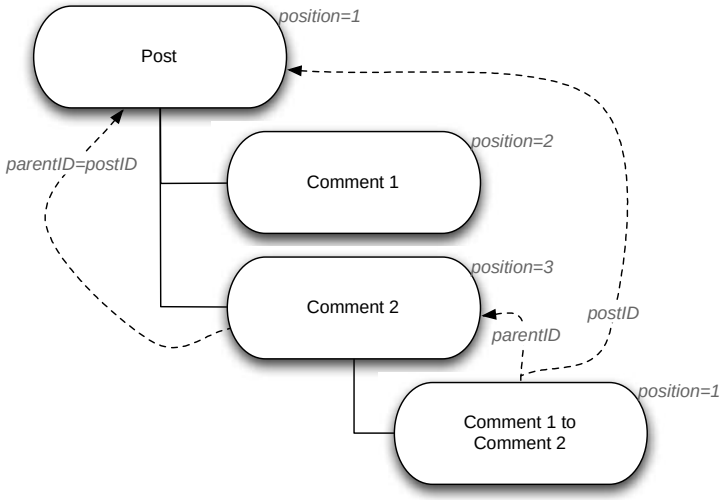

Figure 3: An illustration of the fields parentID, postID and position.

3.2 Data Sources

The list of newspapers selected for data extraction in D2.2 is as follows:

Country	Type	Name	URL
English	News	The Guardian	http://www.theguardian.com/uk
English	News	The Independent	http://www.independent.co.uk/
English	News	The Standar	http://www.standard.co.uk/
France	News	La Provence	http://www.laprovence.com/
France	News	Le Figaro	http://www.lefigaro.fr/
France	News	Le Monde	http://www.lemonde.fr/
France	News	L'Express	http://www.lexpress.fr/
France	News	Les Echos	http://www.lesechos.fr/
France	News	Libération	http://www.liberation.fr/
France	News	20 Minutes	http://www.20minutes.fr/
Italy	News	Corriere della Sera	http://www.corriere.it/
Italy	News	Metronews	http://www.metronews.it/
Italy	News	Il Messaggero	http://www.ilmessaggero.it/
Spain	News	El Mundo	http://www.elmundo.es/
Spain	News	El País	http://elpais.com/
Spain	News	El Períodico	http://www.elperiodico.com/
Spain	News	La Vanguardia	http://www.lavanguardia.com/

Figure 4: list of newspapers selected for data extraction in D2.2

For each of these sources, an example page with comments can be found at http://5.9.95.170/dokuwiki/doku.php?id=sensei_data:data_requirements&#newspapers

These newspapers were queried with the topical query: “Europe OR Europa” and 1000 documents were extracted from each source, resulting in over 3000.
Besides these high-profile newspapers, in order to obtain mentions form general blogs and forums, we issued the query “Europe OR Europa” to Google Search using the news, forums and blogs filters for the time-periods of “this month” and “this year” (the queries were executed in March 2014), resulting in six paginated queries and over 5000 mentions.
Besides news, blogs and forums, we have collected over 20 days all the Tweets, Google+, Youtube and Facebook comments for four major newspapers in three languages (see Figure 5).

A total of 350k items where collected in this manner. This is a live and growing collection.

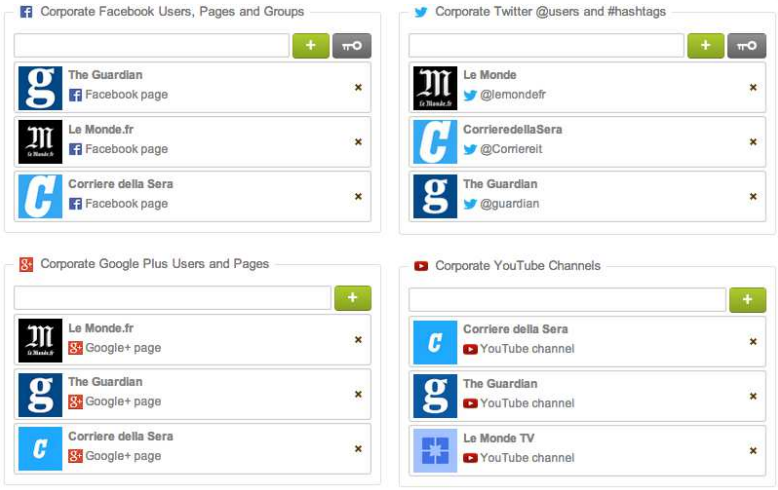

Figure 5: Social media channels for three major newspapers in three languages

3.3 Topics and Uses Cases

The Social Media Use Case Design proposed in D1.1 has been used as a preliminary guideline towards the type of social media content that should be collected. Although the complete data set will be provided in D2.2, we have already tried in D1.1 to collect data useful for some of the different use cases.

Several topics of interest where chosen to track English news to develop the use cases (See Figure 6). Furthermore a generic topic “Europe OR Europa” was added to track content in different languages. Finally, social media of several major newspapers were also tracked (see previous sections).

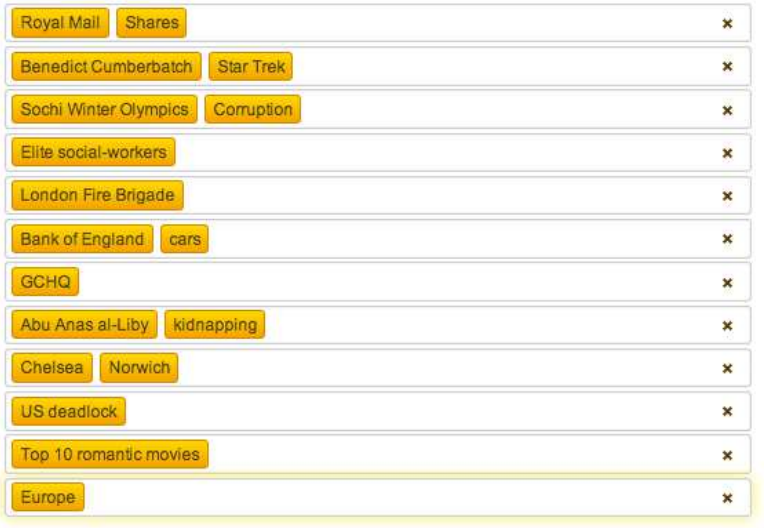

Figure 6: queries chosen to test topic tracking and crawling.

3.4 Parsers and Boiler Plate Detectors

We wrote specialized parsers for each newspaper to convert the newspaper page contents into the Data Schema reported above. In some cases, when newspapers use HTML IFrames to display comments, it was necessary to build distinct parsers for the post and the comments.

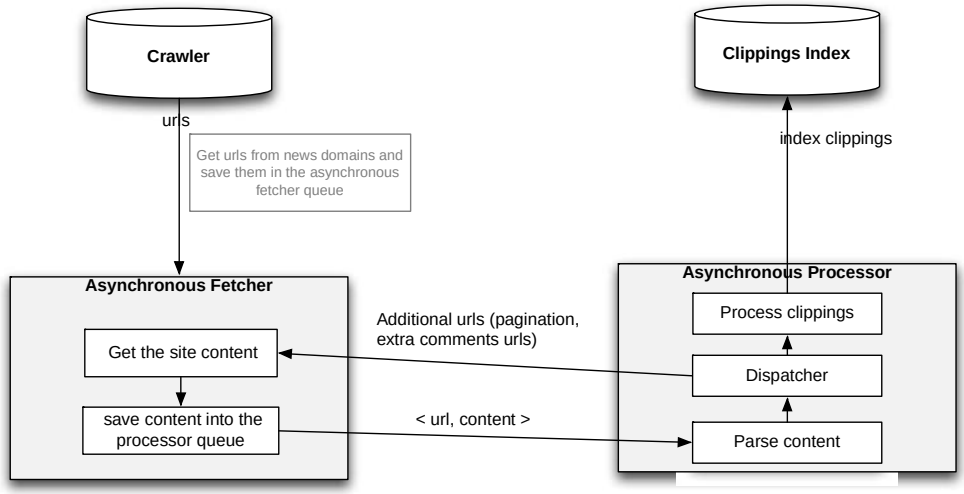

Figure 7: Asynchronous Fetcher developed to serve online fetch requests within an agile parser development environment.

3.5 Crawler Adaptation

Websays provided its high performance crawler, processing and indexing platform. However a number of adaptations were necessary, the main ones being:

- Online (non-batch) crawler: an asynchronous fetcher and processor was developed to allow the crawling of any URL in online mode. An overview of the developed system is sown in Figure 7.
- REST crawl request mechanism: furthermore a REST mechanism has been provided to request the asynchronous crawling of any group of URLs in an online mode.
- Query-agnostic segmenter and parser. Previously Websays only parsed segments of documents matching specific queries. In order to parse and index entire threads and page collections, a query-agnostic segmenter and parser were developed.

3.6 Preprocessing and Evaluation

Preprocessing was done using the standard Websays preprocessing pipeline. This resulted in the following additional annotations on mentions:

- Language
- Named Entities: PERSON, LOCATION, ORGANIZATION;
- Lemmatization;
- Sentiment Classification;

Analysts manually inspected several hundreds of mentions from each of the main newspaper sources as well as many of the blog and forum sources and reported in the parsing and boiler plate detection errors found until all parsers were corrected.

4. Conclusion

The D2.1 deliverable defines the human conversation data sources SENSEI will base its technology design and evaluation. The two source types are collections of speech and social media content. We have summarized the characteristics of the data, the data model, and the process that lead to the final version. In this deliverable we have provided human conversation exemplars and procedures for the acquisition of the entire data set to be delivered in D2.2. The conversation exemplars and the data model will allow the language technology WPs (WP3, WP4, and WP5) to start analysing and processing this data, developing labelling protocols and elaborating the algorithms for the full data set to be provided in D2.2.

ANNEX I: Social Media Data Schema v1

I.1 Description

Data Node	Detailed Data Description	Potential usage for the Use Case
IDs	Integer **externalID** (unique ID of an object in the corpus) Integer **postID** (externalID of the "root" post, e.g. the post originating the conversation) Integer **parentID** (externalID of parent object, e.g. may be a comment of a comment) Integer **versionID** (ID for the version of edited posts) String **sourceID** (URL of the post) String **AuthorID** (...) String **domain** (domain in the url of the post) String **apiObjectID**; (id of the post on the original media platform) String **apiAuthorID**; (the author ID in the native API where this comes from: e.g. Facebook user id) String **apiToAuthorID**; (ID in the native API of the author targeted by the message (eg in a comment))	
Type	enum **superType** {social, news, blog, forum, video, other} enum **type** {post, comment } enum **sourceType** {Facebook, Twitter, news, blog, … } enum **authorType** { anonymous, user, fanpage, group, }	perspective of interests of news
Data	String **title**; (title of the post – if available) String **keywords**; (keywords of the post – if available) String **text**; (body of the post with HTML) String **textHTML**; (body of the post without HTML)	get topic-related interests of readers get topic-related interests of readers
Metadata	POST: Integer **numOfComments**; (number of comments) Integer **numOfLikes**; (Likes in Facebook, News and blogs) Integer **numOfDislikes**; (Dislikes of youtube or metronews) Integer **numOfViews**; (Youtube views) Integer **numOfFavorites**; Twitter Favorites) Integer **numOfReTweets**; (ReTweets) Integer **numOfShares**; (Shares in facebook or g+) USER: String **author**; (author of the post, surface username) String **authorProfilePictureURL** (url of the user profile picture) Integer **user_numOfFollowers**; (Followes of twitter) Integer **user_numOfFollowing**; (Following in Twitter) Integer **user_numOfFriends**; (Friends of facebook and other social) List<String> **authorsMentioned**; (other authors mentioned in the post) List<String> **authorSource**; (from re-tweets, etc.) MISC: enum **mood** (*Indignato*, *Triste*, *Preoccupato*, *Divertito*, *Soddisfatto*, *Uknown*) Integer **moodStrength** (mood strength in corriere.it) String **inReplyTo** (addressee of comments) enum **embeddedMediaType** {text, text&photo, text&video, text&link, photo, video, link} String **tags**; (tags of the post – if available) Boolean **isBestComment** (label for "guardian picks" and best comments) String **pictureURLs** (URL of pictures in articles or posts) String **mediaURLs** (URL of video or other media included in articles or posts)	get temperature on particular issues get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers get topic-related interests of readers *get user's influence* *get user's influence* *get user's influence* get temperature on particular issues get temperature on particular issues
Timestamps	Date **date**;(clipping date real or extracted or guessed)	

	Date **indexingTimeStamp**; (indexing date) Integer **timeRank** (date/time-based ranking for visualization e.g. 1,2,3,4 from the oldest to the newest)	perspective of interests of news
Matches	String **crawlQueryMatch** (text that matched the query that triggered crawling this clipping) Integer **crawlQueryID** (id of queries that triggered crawling this clipping)	get temperature on particular issues get temperature on particular issues
Localization	String **authorLocation**; (geo-location of the author)	

PREPROCESSING:

Annotations	enum **langDetected**; (language of match automatically detected) enum **langReported** (language of the social interface displayed to the user) enum **websaysPolarity**; (sentiment analysis) List<String> **clusters**; (clustering together similar posts based on matchKeywords, matchAuthorNames or Polarity)	get temperature on particular issues have summary to join conversation
Metrics	Integer **count1**; (overall impact (e.g. number of re-tweets))	perspective of interests of news
NLP	String **nlp_chunk**; (shallow chunks annotation) String **nlp_pos**; (shallow part of speech annotation)	

I.1 Schema

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace="http://websays.com/" xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="senseiClipping">
    <xs:complexType>
      <xs:sequence>
        <xs:element type="xs:int" name="externalID"/>
        <xs:element type="xs:int" name="postID"/>
        <xs:element type="xs:int" name="parentID"/>
        <xs:element type="xs:int" name="versionID"/>
        <xs:element type="xs:string" name="sourceID"/>
        <xs:element type="xs:string" name="authorID"/>
        <xs:element type="xs:string" name="domain"/>
        <xs:element type="xs:string" name="apiObjectID"/>
        <xs:element type="xs:string" name="apiAuthorID"/>
        <xs:element type="xs:string" name="apiToAuthorID"/>
        <xs:element type="xs:string" name="superType"/>
        <xs:element type="xs:string" name="type"/>
        <xs:element type="xs:string" name="sourceType"/>
        <xs:element type="xs:string" name="authorType"/>
        <xs:element type="xs:string" name="tittle"/>
        <xs:element type="xs:string" name="keywords" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:string" name="text"/>
        <xs:element type="xs:string" name="textHTML"/>
        <xs:element type="xs:int" name="numOfComments"/>
        <xs:element type="xs:int" name="numOfLikes"/>
        <xs:element type="xs:int" name="numOfDislikes"/>
        <xs:element type="xs:int" name="numOfViews"/>
        <xs:element type="xs:int" name="numOfFavorites"/>
        <xs:element type="xs:int" name="numOfReTweets"/>
        <xs:element type="xs:int" name="numOfShares"/>
        <xs:element type="xs:string" name="author"/>
        <xs:element type="xs:string" name="authorProfilePictureURL"/>
        <xs:element type="xs:int" name="user_numOfFollowers"/>
        <xs:element type="xs:int" name="user_numOfFollowing"/>
        <xs:element type="xs:int" name="user_numOfFriends"/>
        <xs:element type="xs:string" name="authorsMentioned" maxOccurs="unbounded"
minOccurs="0"/>
        <xs:element type="xs:string" name="authorSource" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:string" name="mood"/>
        <xs:element type="xs:int" name="moodStrength"/>
        <xs:element type="xs:string" name="inReplyTo"/>
        <xs:element type="xs:string" name="embeddedMediaType"/>
        <xs:element type="xs:string" name="tags"/>
        <xs:element type="xs:string" name="isBestComment"/>
        <xs:element type="xs:string" name="pictureURLs" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:string" name="mediaURLs" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:dateTime" name="date"/>
        <xs:element type="xs:dateTime" name="indexingTimeStamp"/>
        <xs:element type="xs:int" name="timeRank"/>
        <xs:element type="xs:string" name="crawlQueryMatch"/>
        <xs:element type="xs:int" name="crawlQueryID"/>
        <xs:element type="xs:string" name="authorLocation"/>
        <xs:element type="xs:string" name="langDetected"/>
        <xs:element type="xs:string" name="langReported"/>
        <xs:element type="xs:int" name="websaysPolarity"/>
        <xs:element type="xs:string" name="clusters" maxOccurs="unbounded" minOccurs="0"/>
        <xs:element type="xs:int" name="count1"/>
        <xs:element type="xs:int" name="nlp_chunk"/>
        <xs:element type="xs:int" name="nlp_pos"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

ANNEX II : Example Teleperformance Monitor Detail

Teleperformance CCMS Thursday Jul 11, 2013 16:10:30 CEST

Monitor Detail

Coach Monitor Delete Monitor Edit Monitor

Ident 840911
Name Nome Agente

Monitor Summary

Monitor Ident	26297
Monitor	Sky Ib 100 400 v3
Client	Sky Italia Spa
Target Program	Sky DURAZ Inbound Operations Agents
Reviewer	Quality che ha effettuato l'ascolto
Monitor Date	2013-07-01
Location	TPAL Durazzo
Client	Sky Italia Spa
Program	Sky DURAZ Inbound Operations Agents
Manager	Supervisore dell'operatore
Employee First View	2013-07-01 13:26:44
Employee Last View	2013-07-10 09:10:19
Employee View Count	4

Monitor Score

Score	52.00
Possible Score	65.00
Percent	80.00
Average Score / Question	3.47

Show or Hide Comments

This will allow you to show and/or hide all comments and coaching.

Show All Hide All

Apertura e chiusura della chiamata

	Weight 15.00	Weighted Average 83.33
1. Rispetta Apertura e Chiusura		100.00
2		
2. Intonazione adeguata		100.00
2		

3. Cortesia e disponibilita' 1	66.67

Cortesia ed empatia

	Weight 15.00	Weighted Average 50.00
1. Il tono e le parole utilizzate esprimono cortesia ed apertura all'ascolto 1		83.33
2. Empatico, capace di comprendere appieno lo stato d'animo e la personalità altrui per poi adeguare la propria comunicazione 1		83.33
3. Interagisce costruttivamente attraverso domande di verifica e guida 1		50.00
4. Si assicura della soddisfazione del suo interlocutore 1		66.67

Problem solving

	Weight 20.00	Weighted Average 100.00
1. Individua le risposte più adatte ed efficaci a situazioni difficili attraverso il tentativo di strade diverse 2		100.00

Correttezza contenuti erogati

	Weight 30.00	Weighted Average 100.00
1. Rispetto alla richiesta diretta od indiretta, risponde in modo corretto e completo, con la dovuta ricchezza di dettagli 2		100.00

Informazioni su prodotto e programmazione Sky

	Weight 20.00	Weighted Average 50.00
1. Aggiunge dettagli ed informazioni inerenti il prodotto e la progarmmazione 1		50.00

KPI

	Weight 0.00	Weighted Average N/A
1. Call Reason Altro		88.89
2. One Call solution Si		0.00
3. Durata 18.48 min		
4. TMG 467		
5. Ripetute 30.7%		
6. CSI 92.2%		

Commenti

Weight 0.00 Weighted Average N/A

1. Note Comunicazione / Prodotto

12888711-Il cliente chiama chiedendo assistenza in quanto non riesce a visualizzare i canali dal 109 e in poi. La risorsa informa che il channel pack intrattenimento non fa parte nel suo abbonamento ma il cliente chiede di recuperare il pacchetto completo ma la risorsa informa che provvede con il change di channel pack. Propone lo sconto MGM. Su richiesta del cliente per il Mysky Hd e contestulamente del Multivision , la risorsa crea dei vuoti comunicativi mentre verifica il costo d'adesione. Utilizza lo script di chiusura

Punti di forza: Cortese e gentile verso il cliente.

Punti deboli: Non propone il passaggio al nuovo listino in prima battuta

Suggerimenti: Deve avere praottivita' commerciale e sintetizzare le informazioni fornite

CSI SKY

Weight 0.00 Weighted Average 100.00

1. In riferimento all'ultima chiamata, quanto è soddisfatto del tempo di attesa prima di parlare con un operatore?	
N/A	
2. Pensi ora all'operatore che Le ha risposto, quanto è soddisfatto della sua cortesia?	100.00
9	
3. Quanto è soddisfatto della rapidità dell'operatore nel rispondere alle sue richieste?	100.00
9	
4. Come valuta la competenza dell'operatore nel rispondere alle sue richieste?	100.00
9	
5. Come valuta la capacità dell'operatore di farsi carico delle sue richieste?	100.00
9	
6. Tutto considerato, quanto si ritiene soddisfatto della sua chiamata al servizio clienti SKY?	88.89
9	