

COCKPIT**FP7-248222**

*Citizens Collaboration and Co-Creation in Public
Service Delivery*

Deliverable D2.2**Multilingual Content Analysis Scheme**

Editor(s):	Ilias Spais, Christos Georgousopoulos
Responsible Partner:	ATC, INTRASOFT International S.A.
Status-Version:	Final – v0.4
Date:	30/06/2011
EC Distribution:	Restricted to other programme participants (including the Commission Services)

Project Number:	FP7-248222
Project Title:	COCKPIT

Title of Deliverable:	Multilingual Content Analysis Scheme
Date of Delivery to the EC:	30/06/2011

Workpackage responsible for the Deliverable:	WP2 – Citizens’ Opinion Mining and Deliberation
Editor(s):	Ilias Spais (ATC), Christos Georgousopoulos (INTRASOFT)
Contributor(s):	Ivan Ficano (ENG), Federica Bettio (Venice), Panagiotis Kokkinakos (NTUA), Vassilis Michalitsis (YPES), M.T. Smits (CiTilburg), Yéhia TAHER (UnTilburg)
Reviewer(s):	Nidhi Rajshree (IBM)
Approved by:	All Partners

Abstract:	This report forms a complementary documentation to the source code files that are delivered as part of this Deliverable which is of type ‘Prototype’. It briefly describes the schemas of which the COCKPIT Opinion Mining module is equipped with, to identify Web 2.0 sources of content relevant to the selected public services that will be piloted.
Keyword List:	content schemas, opinion mining, ontology, Web 2.0 sources, url list, crawler, crawled links

Document Description

Document Revision History

<i>Version</i>	<i>Date</i>	<i>Modifications Introduced</i>	
		<i>Modification Reason</i>	<i>Modified by</i>
v0.1	17/06/2011	Documentation of the report that is accompanied to the source code files supplied in the context of this deliverable that is of type 'Prototype'.	ATC
v0.2	28/06/2011	Delivery of final version of source code files.	INTRASOFT, ATC
v0.3	28/06/2011	Internal review of documentation and verification of supplied source code files. Provision of comments	IBM
v0.4	30/06/2011	Preparation of final version. Address of reviewing comments and final formatting.	INTRASOFT

Table of Contents

EXECUTIVE SUMMARY	5
1 FILES THAT ARE DELIVERED	6
1.1 GREEK PILOT CASE	6
1.2 ITALIAN PILOT CASE	7
1.3 DUTCH PILOT CASE	7
1.4 CORE MODULE IDENTIFICATION SCHEMA	7

Table of Figures

FIGURE 1: ONTOLOGY CREATION GUIDELINES	6
--	---

Executive Summary

This is a supplementary documentation that describes all the material needed in order to produce the schemas of which the COCKPIT Opinion Mining (OM) module is equipped with, to identify Web 2.0 sources of content relevant to the selected public services that will be piloted.

In order to determine the citizens' sentiments about a specific public service delivery, relevant content from user generated sources need to be first identified and collected, and then appropriately analysed to produce the citizens' sentiments. Identification and collection of relevant content is very crucial in COCKPIT's opinion mining methodology and that is why a demo case was organised before the production of the actual schemas. Results and lessons learned from this demo were documented in deliverable D2.2.0 "Opinion-mining demonstration in Web 2.0 approach" and made the basis for further guidance and directions in the work that was provided in the deployment of pilot use cases.

According to the project's work plan, the COCKPIT OM module will be utilised in three different pilot cases. The selected public services set the basis for the identification schemas, and the respective partners followed the guidelines of D2.2.0 to produce the requested material.

The schema of each piloting public service is composed of the following information expressed into the local language:

- a generic public service ontology consisting of general terms related to policy, government, public services etc., and
- a specific public service ontology pertinent to the domain of the respective pilot case scenario.

The guidelines that describe the creation process of the associated identification schemas, as illustrated in figure 1 may be found in Deliverable D2.2.0, where the use of those schemas within the OM process is provided in Deliverable D2.1.1.

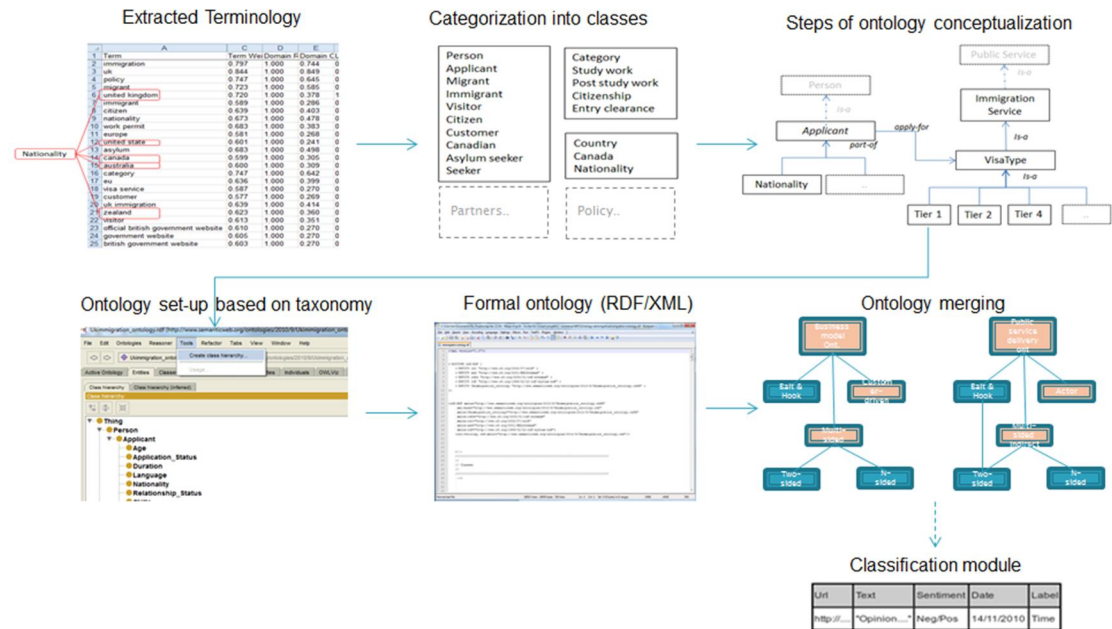


Figure 1: Ontology creation guidelines

1 Files that are delivered

A description on the files associated with the identification schema of each pilot case is provided in the following sub-sections.

1.1 Greek pilot case

- Public service selected:
 - GR01 - Access, extracts of insurance record in Social Security Organization.
- Multi-lingual Content Analysis Scheme:
 - IKA-GR.owl. Ontology used by the OM in Greek language.
 - IKA-EN.owl. Translation of the Greek ontology to English language.
 - IKA-OM-schemabased-sources.xlsx. List of sources that are crawled from COCKPIT's Web crawler.
- All the files are included in the directory:
 - Greek pilot case\content analysis schema

1.2 Italian pilot case

- Public service selected:
 - *IT04 - Integration between Iris and GeoLP.*
- Multi-lingual Content Analysis Scheme:
 - *IRIS-IT.owl.* Ontology used by the OM in Italian language.
 - *IRIS -EN.owl.* Translation of the Italian ontology to English language.
 - *IRIS-OM-schemabased-sources.xlsx.* List of sources that are crawled from COCKPIT's Web crawler.
- All the files are included in the directory:
 - *\Italian pilot case\content analysis schema*

1.3 Dutch pilot case

- Public service selected:
 - *NL05 - Werkplein Job intermediation.*
- Multi-lingual Content Analysis Scheme:
 - *Werkplein-NL.txt.* Hierarchical taxonomy of relevant terms.
 - *Werkplein-EN.txt.* Hierarchical taxonomy of relevant terms translated to English language.
 - *Werkplein-OM-schemabased-sources.xlsx.* List of sources that are crawled from COCKPIT's Web crawler.
- All the files are included in the directory:
 - *\Dutch pilot case\content analysis schema*

1.4 Core module identification schema

- Generic Taxonomy:
 - *GenericPublicServiceSimplifiedExtractedTaxonomy.txt.* Generic taxonomy containing generic terms on policy, government, public services etc.
 - *PublicServiceModel.ecore.* Latest version of the meta-model (ecore file).

- All the files are included in the directory:
 - *\Generic Ontology and metamodel*