PROGRESS REPORT

Grant Agreement number: 250467

Project acronym: ATLAS

Project title: **Applied Technology for Language-Aided CMS**

Project type: ☐ **Pilot A** **V Pilot B** ☐ **TN** ☐ **BPN**

Periodic report: **1st ☐ 2nd ☐ 3rd V 4th ☐**

Period covered: **from 01.03.2011 to 31.08.2011**

Project coordinator name, title and organisation:

Anelia Belogay, CEO, Diman Karagiozov, CTO,

Tetracom Interactive Solutions

Tel: 35924950444

Fax: 35924950443

E-mail:anelia@tetracom.com, diman@tetracom.com

Project website address: www.atlasproject.eu

DECLARATION BY THE PROJECT COORDINATOR

I, as coordinator of this project and in line with my obligations as stated in Article II.2 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

- The project (tick as appropriate):

  V has fully achieved its objectives for the period;
  has achieved most of its objectives for the period with relatively minor deviations;
  has failed to achieve critical objectives and/or is deviating significantly from the schedule.

- The public Website is up to date;

- [this point only applies to projects with actual cost reimbursement] To my best knowledge, the information contained in the financial statement(s) submitted as part of this report is in line with the actual work carried out and consistent with the reported resources and if applicable with the certificates on financial statements.

Name and position of Coordinator: Anelia Belogay, CEO

Date: 23.10.2011

Signature:

PUBLISHABLE SUMMARY

Introduction

The advent of the Web revolutionized the way in which content is manipulated and delivered. As a result, digital content in various languages has become widely available on the Internet and its sheer volume and language diversity have presented an opportunity for embracing new methods and tools for content creation and distribution. Although significant improvements have been made lately in the field of web content management, there is still a growing demand for online content services that incorporate language-based technology. Mechanisms such as automatic annotation of important words, phrases and names, text summarization and categorization, and computer-aided translation could facilitate the process of manipulating heterogeneous multilingual content as well as enhance end-user experience by allowing for better content navigation. This project unifies such mechanisms in a common software platform called ATLAS and builds three separate solutions around this platform.

Summary description of project objectives

The consortium will adjust and integrate several existing software components, assembling a platform for multilingual web content management called ATLAS, and a visualization layer called i-Publisher, which adds to the platform a powerful web-based point-and-click tool for building, reusing and managing multilingual content-driven web sites. An instance of i-Publisher will be made publicly available as an online service. i-Publisher will also be used to build two thematic content-driven web sites – i-Librarian and EUDocLib.

The ATLAS project aims to meet the following objectives:

Software platform and services, demonstrating the latest achievements in the field of multilingual web content management and addressing the needs of individuals and organizations for easier web site building and content publishing.

- Liaison with the Europeana and EuroMatrix Plus initiatives in order to foster language diversity in content creation and distribution

- Interoperability by conforming to a number of widely recognized web, natural language processing, and content management standards

- Sustainable management format to ensure the progress of the project

- Mechanisms and procedures that enable and simplify the addition of new languages to the ATLAS platform, thus targeting all major European languages after the successful completion of the project.

Expected final results

The primary goal of the ATLAS project is to facilitate organizations and individuals who manage and publish multilingual content. Thus, the project solutions will not merely meet the needs of modern multilingual content management, but also create value for all users.

Main expected final results:

- The software solutions built during the project reveal the true value, capabilities and power of several existing tools for web content management, multilingual versioning, and natural language processing by combining them in an innovative manner and offering the end results to the general public at no cost.

- With i-Librarian and i-Publisher users can easily create, manage and publish multilingual content without installing and maintaining a standalone system. Nevertheless, they retain full control over their content regardless of whether it is in their private, shared or published workspace. EUDocLib provides easy and intuitive access to a vast collection of EU law documents.

- The ATLAS platform is designed with extensibility in mind, which allows for easy addition of tools for currently unsupported languages as well as new tools for already supported languages.

- Furthermore, ATLAS significantly reduces the time and efforts for content authoring and editing because it automatically categorizes, summarizes, annotates and translates documents regardless of their language and format. The software platform enables i-Librarian users to find the most essential texts from large document collections by displaying text summaries and extracted important phrases, words and names.

- Finally, ATLAS improves content navigation by interlinking content items based on text annotations and by automatically placing the content items in appropriate subject categories.

Potential impact

The project brings together advanced technologies for multilingual web content management and text mining (such as automatic annotation, mark-up and translation) in a united platform. The intended software-as-a-service architecture of the envisaged solutions, which demonstrate the capabilities of the ATLAS platform, and the open-source license, will facilitate the spread of the project output.

Main expected impacts:

- Technological
    - Integration of text mining tools into content management systems
    - Integration of text mining services
    - Stable and more efficient Machine Translation modules for the project languages. The language pairs considered in ATLAS are covered by Google Translation but with very low quality. On the other hand, these language pairs have strong relevance for the Central- and East-European commercial space.
    - Contribution to the development of text processing chains for languages, which lack resources at present
    - Adherence to and promotion of existing and future web standards

    - Practical and economically viable solutions for nearly-automatic provision of multilingual online content and services for some EU languages
- Social
    - Facilitate exchange of information and knowledge
    - Simplify authoring, management and exploitation of heterogeneous multilingual content
    - Address the needs of a large number of people belonging to different target user groups – individuals and organizations
    - Cross the language barrier
    - Facilitate culture exchange
    - Liaise with Europeana and EuroMatrix Plus – The liaison with EuroMatrix Plus will be established at the beginning of the project. Europeana will be approached by the end of the first year, when the consortium will be able to demonstrate the potential value of ATLAS to the European digital library.

Use

The ATLAS platform as a whole and also some of its standalone components are beneficial to different groups of users. Thus the consortium has distributed the potential users of each major software component into several target groups while paying special attention to the needs and requirements of each group. The table below summarizes this distribution:

Target groups

Component	Target group
ATLAS (includes KMS Content Management System, Text Mining engine, Search engine, Machine Translation engine) + i-Publisher (ATLAS web-based graphical user interface for building interactive, content-driven web sites)	Web design companies – faster prototyping, web design and site building
	Hosting companies – as part of hosting packages
	Education, Media, Publishing, Non-profit, Government
Text Mining engine	Online bookstores
	Digital libraries/repositories
	News agencies/websites
i-Publisher (as online public service)	Small enterprises
	Non-profit organizations
i-Librarian (thematic content-driven web site built with i-Publisher)	Students, Researchers
	Readers
EUDocLib (thematic content-driven web site built with i-Publisher)	The general public

Table 1: Target groups

More information including project details, news, and contact information can be found at:

www.atlasproject.eu

PROJECT PROGRESS

Project objectives for the period

With regard to the management objectives set for the reported period the following tasks have been completed:

- The management and coordination framework ensured the smooth progress of the project.
- The control of the quality and timely delivery of project deliverables as well as monitoring of the allocation and distribution of the project resources resulted in implementation of all tasks as planned.
- One project meeting was organized in Luxembourg in June. The Consortium made an overview of the project progress, discussed the project presentation for the Review meeting and defined the next steps needed to achieve the objectives for the next period.
- The leader of WP 3 restricted considerably the involvement in the project and the Coordinator replaced the partner and took the responsibility for the package.
- An amendment was prepared by the Coordinator. It included the updated version of DoW reflecting the cancelation of Croatian as a project language and the replacement of the partner responsible for the categorization.
- The first year report covering month one through month twelve of the project was prepared and submitted to the EC.
- The first review meeting was held in Luxembourg. The two deliverables, requested by the Commission for resubmission, were resubmitted by the Coordinator.

The work done in terms of the technical objectives set for the period includes the following:

- Improvements in the linguistic platform were done – a new noun phrase extractor was implemented together with new name entity extractor for English. In addition, a new functionality to process the text in paragraphs and not as a whole text was implemented and as a result the user gets intermediate results from language processing almost immediately after uploading a file.
- Improvements of usability of i-Publisher – export/import and reuse of a web site was implemented; various functionalities were extended in order to increase the productivity of the users.
- **i-Publisher Simple Mode** was built and integrated into ATLAS CMS. The requirement was defined by the test users involved in the first test round held in March. The Simple Mode is a simplified layer of i-Publisher providing the inexperienced users with ready-to-be-used thematic web sites and themes.
- The first drafts of the user guides of i-Publisher Advanced Mode and i-Librarian are assessable. The complete version will be delivered in PM 22.
- The implementation of the technical documentation of the three services has started. The final version will be delivered in PM 22.
- The existing categorization engine was extended to support several algorithms, namely Relative Entropy, Naive Baysean, Class-Featured Centroid, etc. In addition, the engine now supports combining of different classifiers which results in better predictive performance**.**
- An update of the core annotation model with new features (head noun annotation, co-reference chains annotations), considered vital for higher-level text processing tasks or visualization, was done by Tetracom.

- The implementation of the language processing chains (LPCs) for all project languages was carried out, targeted at providing tools for sentence-splitting, tokenization, lemmatization, POS tagging, NP extraction and NE recognition.
- The LPCs were prepared for integration. Bulgarian, Polish and Greek are ready to be integrated and German and Romanian are in a test phase.
- The internal verification of the chained NLP tools was carried out.
- The testing procedures were verified and the draft internal documentation was updated accordingly (with internal test report templates, description of the test infrastructure and test result templates)
- Two different summarization approaches that will provide the user with a summary depending on text length were defined. The partners have started collecting the text corpus for short summaries per every language.
- Two long-text summarization methods were developed: a “shallow” one and another that exploits the full chain of summarization tools – Language Processing Chain, RARE (Robust Anaphora Resolution Engine), Discourse Parser, clause splitter and summarizer.
- A summarisation chain is implemented for English and Romanian; for some of the other languages (Bulgaria, Greek, Polish) training corpora and experts summaries are under development.
- RARE tool has been updated. It includes now a genetic algorithm able to learn rule weights and priorities, as well as window values from a gold annotated corpus.
- The 1st round of the ATLAS system User Acceptance Evaluation has been conducted and the results were presented to the Consortium and the Commission. As a result, additional simplified mode was added to the functionality of i-Publisher, the quality of the work of the English chain was improved.
- The MLeCeL LL was established in Sofia and a website is accessible at http://livinglab.itd-bg.eu/ . The two host organization teams set their laboratories and started conducting experiments with the services.
- Technical Indicators for the CMS, summarization, machine translation, Cross-lingual Content Retrieval have been drafted.

Work carried out on dissemination for this period includes:

- The new project website was created with i-Publisher.
- Preparation for several major dissemination events:
    - ATLAS Workshop “Integration of multilingual resources and tools in Web applications workshop” @ Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011)
    - ATLAS – Multilingual Language Processing Platform paper presentation at SEPLN 2011
    - The ATLAS project and i-Publisher, i-Librarian will be presented and demonstrated in a stand at WEBIT 2011 (webit.bg) to a wider audience (5000 visitors, 40 countries)
- A new vision of the project and its services is applied to the dissemination materials and services websites.
- The project has been presented to linguists, computational linguists, NLP engineers, IT scientists, key players in Language Resources and Technologies and representatives of social sciences and

humanities, business people, students and researchers in conferences, meetings described in details in WP 8.

Main results for the reported period

- Relaunch of the project web site. The website is now exclusively supported by i-Publisher.
- Significant improvements of the usability and quality of the services.
- Extensions of the Categorisation tool with several new algorithms and a "voting" system for improving the automatic categorisation results.
- Alfa aversion of all language chains are integrated in Atlas

Work progress and achievements during the period

Work package 1 - Work package 1– Project Management

The package description is placed in chapter "Project management".

Work package 2 – Software Specification, Implementation and Deployment

The work package focused on preparing the drafts of technical documentation and user guides for the ATLAS platform and the online services. In addition, Atlas, i-Publisher, i-Librarian and EUDocLib were improved by updates of tools, modules, and additional implementations in order to provide to the end users qualified services.

T2.5 (23 PM) – i-Publisher – implementation and deployment

The launch of the three services was followed by an evaluation done by users within the first test round. The test groups of 33 users assessed the technical performance, the level of fulfillment of the system specifications and the level of user expectations fulfillment and provided Consortium with their feedback. As a result several major improvements were defined, discussed with the Commission and Tetracom started with their implementation:

- I-Publisher Simple Mode. The complexity of i-Publisher, a powerful instrument for creating complex web sites, made it difficult for the group of inexperienced test users to work with it. As it is a wider and an important target group, Tetracom decided to extend the web based tool i-Publisher with simplified layer for non experienced users. The users will choose from ready-to-be-used websites or they will create websites with predefined visualisation (themes) still having the possibility to define the content and its structure. As none of the excising now online services, some of the ready-to use websites will provide the functionalities related with the linguistic framework so that the user can benefit from its applications like automatic annotations, automatic translation.
- New English noun phrase extractor. The Open NLP noun extractor, initially integrated in the English NLP, was replaced by a new one. The extractor is based on the theory of finite-state language processing and contains set of language-specific rules for recognition of noun phrases described by an English linguist. The NP extractor can be used for other languages as well.
- Improvement of name entity recognizer. Several updates in the architecture of Atlas were needed in order for the JRC Names Library to be integrated into Atlas and as a result we provide the user with name recognition covering wider set of person names and organizations. The JRC Names UIMA wrapper is now used in other project Language Processing Chains.

- Stability and performance optimizations. I-Publisher and the underlying language processing chains framework were optimized in terms of performance by integration of caching components on different levels of the system architecture. The web sites, powered by ATLAS, were tested with the Apache HTTP server benchmarking tool.

T 2.6 - i-Librarian – implementation and deployment

i-Librarian is a thematic web site (online service), which encourages visitors to register and get a personal workspace where they can store, share and publish various types of documents and have them automatically categorized into appropriate subject categories, summarized and annotated with important words, phrases and names.

- The noun phrase extraction and the name entity extraction in English were significantly improved by using the updated versions of the noun extractors and name entities service.
- A new categorization tree containing 86 categories grouped in main groups enhances the users in cataloging their documents automatically.
- The "public" library contains a collection of more than 22'000 books in English from Project Gutenberg.

T 2.7 – EUDocLib - implementation and deployment

EUDocLib is a publicly accessible repository of EU law documents from the EUR-LEX collection. This web site (online service) provides enhanced navigation and easier access to relevant documents in the user's language.

- All documents of EUR-LEX collection, available in English, added up to the 1 January 2011, have been imported into ATLAS (resp. into EUDocLib).
- The categorization methods applied to the EUR-LEX collection did not give the initially expected results due to the fact that the trained model is based on noisy data. The harvested data will be now clean up and another classification approach will be taken.

T 2.8 - Technical documentation

Tetracom started writing detailed technical documents describing the ATLAS platform and the online services. Documentation will be made publicly accessible on the project web site. ATLAS technical documentation draft, i-Publisher technical documentation draft and i-Librarian and EUDocLib technical documentation drafts include a thorough description of the platform architecture and design, and emphasize the multilingual capabilities of the software. The final version will be ready in PM 22. The delay is caused by the additional implementation of i-Publisher Simple Mode which affected Atlas architecture and the relation between the software components.

T2.9 – User guides

Tetracom started preparing the user guide system for i-Publisher. The comprehensive hint system is ingrained in i-Publisher Advanced Mode. In addition, video films enhanced the user in perceiving the main complex workflows.

The major i-Publisher workflows will be described in the "Atlas quick user guides" as well and will be ready in PM 22.

Work package 3 – Automatic Categorization

The focus of WP3 is implementing a tool that categorizes texts in the target languages and can be tuned to heterogeneous domains. The ultimate goal of the automatic categorization is a qualitatively better and more effective organization of content in ATLAS and respectively in the online services.
The working package leader stepped out of the project and the Tetracom replaced UniZd responsibility within this package.

T3.1 Prototype and integration

- UniZD provided a categorization tool (Python module) that supports three automatic categorization algorithms – Naive Bayesian, Relative Entropy and TFIDF score.
- The UniZD tool was successfully integrated into the ATLAS framework and an English model was built.
- Tetracom conducted several experiments to study the categorization performance in relation to the type of features used for vector space generation. The following features types were used – token based, lemma based, noun phrases based, head token based, and combinations of the former.
- Tetracom extended the ATLAS categorization framework with the following changes:
    - Tetracom implemented an additional categorization algorithm - Class-Feature-Centroid classifier.
    - The categorization framework now supports several instances of categorization algorithms, each instantiated with different feature type.
    - Support for ensemble of categorization algorithms. Multiple models can be used to obtain better predictive performance.
- Tetracom built an English model for 80 categories, which is now accessible on the site (‘Topics’ categorization tree) of i- librarian

T3.2 (25 PM) – Fine-tuning to specific languages

- This task is ongoing. The Bulgarian partner ICL DCL kindly provided Training and Test documents for the 80 categories. Tetracom is currently processing the data and building a Bulgarian model.

T3.3 (18 PM) – Software adjustments (in resp. to technical reviews)

- This task is ongoing. Tetracom implemented part of the UniZD Python tool algorithms in Java in order to assess the benefits of Python performance in contrast to Java maintenance easiness.

T3.4 (14 PM) – Documentation

- This task is ongoing.

Work package 4 – Language Processing Chains

The main focus on WP 4 for the reported period was to:

- update the annotation model with new features that were considered vital for higher-level text processing tasks or visualization,
- carry out the implementation of the language processing chains (LPCs) for all project languages, targeted at providing tools for sentence-splitting, tokenization, lemmatization, POS tagging, NP extraction and NE recognition,
- prepare the LPCs for integration with the test environment,
- carry out internal verification of the chained tools,

- verify the testing procedures and update the draft internal documentation accordingly (with internal test report templates, description of the test infrastructure and test result templates).

The deliverable D 4.1 has its draft version.

T4.1 – Implementation of LPCs for the project languages

Bulgarian

Tokenizer, sentence splitter, NP chunker, POS tagger have been implemented and verified. First versions of the WSD web service and Named Entity Recognizer have also been provided.

Croatian

It was agreed with the project officer that the Croatian tools will not be implemented/integrated in ATLAS.

English

Rules-based NE recognition was tested against the OpenNLP suite to improve performance.

German

UHH adapted the German OpenNLP to be used with the ATLAS framework.

Greek

The following set of components for Greek LPC have been prepared:

Primitive engines of the GR NLP tools

- UIMA compliant primitive engines created for sentence splitter, for tokenizer, for POS Tagger, for NER and the Lemmatizer.

All GR primitive engines were revised according to the recently updated UIMA types.

GR Named Entity Recognizer

Designed and developed as UIMA compatible component. Supports name (person, organisation), location, date, time, money and percentage. Additional support for recognition using regular expressions.

Verification testing with input texts from various domains (reliability and performance analysis).

GR Lemmatizer enhancement

Added more lemmas found in dictionaries.

Added support for MySQL (back-end repository) and for multithreading in order to minimize the time needed to lemmatize a sequence of tokens.

Enhanced implementation by incorporating a cache approach for the lemmas included in the database (using the EhCache open-source project). Work in progress.

GR POS Tagger improvement

Improvements were made in the performance (loading of objects and necessary files for classification) of the tagger.

Improved version of GR Noun Phrase extractor

Morphological rules have been enhanced (more than 6 now are available and tested).

UIMA classes revised to be able to support multithreading if needed. In progress is the migration to the latest available version of the Spejd tool.

GR Anaphora Resolution

A first version became available, based on the RARE tool provided by UAIC. Needs testing and verification.

Polish

In months 13-18 ICS PAS was carrying out implementation, reimplementation, improvement and integration of existing tools from the Polish LPC, namely:

- Spejd-based multiword NP lemmatizer has been implemented from scratch by Łukasz Degórski and later described in his paper presented at Security & Intelligent Information Systems 2011 conference in Warsaw (http://siis.ipipan.waw.pl, 13-14 June 2011). NP lemmatizer was evaluated on a manually annotated small corpus of multiword expressions.
- Existing Named Entity Recognition by Jakub Waszczuk was improved to reflect ATLAS NE model (by supplementing it with ATLAS types not available in the original version – Money, Percentage) and integrated into the test installation.
- Spejd shallow parser used for NP recognition was reimplemented in C++ to improve its efficiency.

The new components of the Polish LPC were made available in the new version of WebCASDebugger. Bug testing has been carried out and library incompatibilities resolved.
Notes on the LPC test platform have been sent to Diman Karagiozov along with internal test report template and initial D4.1 version concentrating on documenting the test infrastructure (section 4.6) for both pre- and post-integration actions, providing the test result templates (section 4.7) and including the real test results (sections 4.8) in the form of filled in templates from section 4.7.

Romanian

UIAC was carrying out integration of the POS tagger (a tokenizer, a sentence splitter, the POS tagger and a lemmatizer). The NP chunker was being rebuilt due to problems with complicated NPs in its previous NOOJ-based version.

German

Defining the structure of the components in UIMA and adaptation of sevral tools included in the WCDG-System (http://nats-www.informatik.uni-hamburg.de/view/CDG) in order to be used within ATLAS as follows:

Tokeniser

- Primitive Engine,
- includes the Perl Tokeniser from WCDG
- Changes: adapted to deliver positions of tokens, instead of simply listing them

PoS Tagger

- Aggregate Engine
- Wrapper aound the TnT Tagger → here all possible PoS together with their probabilties are generated
- WCDG –Aggregate Engine -> parser. Together with the parsing result the PoS is disambiguated

Lemmatiser

- Aggregate Engine
- Morphological analyser (perl) + disambiguation through WCDG

NP Chunker

- Aggregate Engine
- WCDG + Primitive Engine : Tool for composing NP –Chunks from dependency structures.

General

JRC-Names service provided by the Joint Research Center is now wrapped as UIMA component and is ready for integration within all project Language Processing Chains.

T4.2 – Software adjustments (in resp. to technical reviews)

Tetracom adjusted the ATLAS language processing architecture in respect to the performance and stability. Minor functional adjustments to i-Librarian linguistic functionalities has been made. Atlantis implemented performance optimizations for the Greek Lemmatizer

T 4.3 - ATLAS LPC support extension

Tetracom extended the core annotation schema according to the requirements from the parners. Namely, the NP supports a lemmatized vesion, NE annotation has a provider and common identifier. All UIMA wrapper classes has been migrated to use the AbstractPrimitiveEngine class and the new integration infrastructure.

T4.4 – Integration of LPCs in ATLAS

The English, Bulgarian, Polish and Greek NLP were integrated in Atlas. The Romanian NLP is ready to be integrated (the chains were integrated in the beginning of September) and the German NLP is in a test phase.

T4.5 (18 PM) – Documentation

It is an ongoing task.

Work package 5 - Text Summarization

T5.1 – Implementation of text summarization tools

RARE (Robust Anaphora Resolution Engine) has been updated. It includes now a genetic algorithm able to learn rule weights and priorities, as well as window values from a gold annotated corpus. The expected output is a model that optimises the applications of the set of defined rules.

As explained to the partners in our preceding meetings, different summarisation methods should be employed for short and for long texts.

The short texts summarisation task depends heavily on a set of manually annotated texts for summaries. The value of a summary should be decided by comparing the automated generated one against the summary considered as gold (the comparing metrics is a topic to be decided also). A series of small texts have been chosen for the summarization task. We decided to use short stories that can be easily found in all languages of the project. This way, if this will be decided later as useful, we could organise a comparison among the summarisation tools of the partners.

The following short stories (each less than 4 pages) have been announced to the partners as recommended:

- “Allan Edgar Poe - The Masque of the Red Death”,
- “Arthur C Clarke - If I forget thee, oh Earth”,
- “Grimm Brothers - Sleeping Beauty”,
- “Grimm Brothers - The Queen Bee”,
- “H. P. Lovecraft – Polaris”,
- “H. P. Lovecraft - The other gods”,
- “H. P. Lovecraft - The quest of Iranon”
- “Washington Irving - The adventure of the German student”.

The Romanian and English versions of these short stories have been included in the corpus.

For long texts, we prepare two summarisation methods (besides the openNLP version, available for English and which has already been used by Tetracom): one method can be considered as shallow, as it relies on surface heuristics, mainly language independent, such that each sentence is given a score. Firstly scores are given for each word based on heuristics and then for each sentence the scores of the words are added, and the final score is divided by the count of the words in the sentence. Heuristics used for scoring the words are: the first appearance of a word has a bigger score; if the word is starting with a capital letter its score is increased; the number of appearances of the word also affects the score (rare words have big scores, common words have small or no scores).

For English, based on RACAI, UAIC and openNLP tools, we have developed a summarisation chain, which includes the following processing modules: tokenisation, POS-tagging, NP-chunking (all openNLP brand), lemma (RACAI), and RARE, Discourse Parser, clause splitter and summarizer (UAIC). Using this chain we attained a summary for a text, which has been uploaded on the ATLAS SVN server. The same text has been manually splitted into clauses, and a manually (gold) summary has been provided for comparison. The annotation was done using a list of cue-phrases and with the help of a annotation tool PALinkA we annotated the corpus for clauses boundaries and cue-phrases, according to Rhetorical Structure Theory (2).

A clause segmenter is now under development. It is based on heuristics and patterns learned from the corpus, which will segment sentences into clauses. The marker pattern represents a window of n POS (part-of-speech) to the left and n to the right of the marker, where after several tests we approximate that n will be 3. In the cases where between two clauses there is no marker, the verb pattern will represent a window between the last verbal predicate from the first clause and the first verbal predicate from the second one, which will include the clauses boundary. Based on this patterns, considering distances, scores and heuristics, we will split sentences into clauses.

DCL - IBL progress:
A corpus designed for annotation of anaphora chains has been compiled (app. 100 000 words). The corpus is preprocessed and annotated for tokens, parts of speech, sentence markers, and clause boundaries. The Chooser, the DCL annotation tool, has been redesigned for the anaphoric chains annotation and the hand-made summaries attachment to the texts.

ATLANTIS progress:
GR Anaphora Resolution: A first version became available, based on the RARE tool provided by UAIC. Needs testing and verification.
Compatibility conversion of UIMA output: Conversion of the annotation provided by the Greek UIMA LPC to the XML input needed by the Summarization module. Work in progress.
GR Discourse Markers: Compilation of a list of GR discourse markers.
Training corpora: production of a GR training corpus (clause annotation, discourse markers) for summarisation. Work in progress.

ICS PAS progress:
ICS PAS investigated the prototype of the external Polish summarization system implemented by Joanna Świetlicka and started implementation of the summarization Web service to verify the approach and create baseline environment for future summarization tests. Similarly, as a starting point, ICS PAS continued experiments with Polish anaphora resolution using RARE tool.
ICS PAS implemented a rule-based coreference resolution tool and integrated it with summarization Web service (http://chopin.ipipan.waw.pl/multiservice) to be used in further text summarization-related experiments.
Maciej Ogrodniczuk and Mateusz Kopeć described the end-to-end rule-based coreference resolver for Polish in a paper submitted to DAARC conference (http://daarc2011.clul.ul.pt) and accepted in mid-July.

Work package 6 – Machine Translation and Cross-Lingual Search

1. Machine Translation

The main objective is to provide a machine translation engine, able to provide reasonable translation for assimilation for all language pairs involved in the project.

During the Months 12-18 work was done on 3 topics all involving Task T 6.1

- Defining a workflow for the translation engine
- Study about impact of text genre within the same domain on translqation quality
- Training translation models.

I . Workflow for the MT-Engine
Following the work in Months 1-12, the workflow for the MT-engine was defined

- Hybrid System : Example -based Machine Translation (EBMT) + statistical based MT (SMT).The principle is the following:
    - Prerequisites: each translation request should be accompanied about an information about the domain to which the input belongs (i.e.document was before categorised); and an information

about the period when the document was written (metadata extraction)

  - Input from older than 1850 will not be translated, and a message will be shown to the user
  - Input from Documents between 1850- 1950 will be sent futher to the translation module, however the user will be supported with information that the translation may be not accurate
  - If for the respective domain and language pair, there is a trained model /translation database the translation model is automatic initiated . If not the user will be informed and the model available for the closest node in the categorisation tree (going up in the hierarchy) will be selected
- the input is first processed by EBMT engine. If the input is entirely found in the translation database (which can often happen in case of restricted domains or keywords, short chunks) then the translation equivalent is provided as result.
- in all other cases the input is sent to the SMT engine, which will provide the result.
- the user is asked if he wants to correct the input and/or store for further use

II. Impact of text genre on the translation quality

As the biggest available corpora for a series of language pairs are Europarl and JRC-acquis, which more or less belong to the same domain, we investigated to which extent one corpus may be prefered to the other:

Europarl: collections of EU-Parliament discourses- spoken -like style
JRC-Acquis : written EU- laws, regulations

As for all language pairs the only available common corpus is JRC-Acquis, we investigated to which extent we can:

- replace JRC-Translation Models with europarl-ones for the language pairs for which this is available
- use JRC-acquis translation model also for Europarl-like input.

The results, which will be published in a paper at LTC 2011, Poznan, show that the 2 corpora are complementary, there is very low overlap with respect to the number of common sentences , and also the number of non-common words is quite high.

The conclusion of this experiment lead us to the conclusion that there are 2 options :

- training 2 translation models one on JRC-acquis and one on Europarl. In this case however the system has to be informed through an additional parameter which model to select
- concatenate the 2 corpora. The effect on the translation is currently analysed

III. Training Models and Preparing translation databases for EBMT and Moses SMT

We selected mainly for this task the 3 big available parallel Corpora for most of the language pairs: JRC-Acquis, Europarl and SETimes.

The JRC-Acquis was extended with a small corpus on EU-constitution available under http://opus.lingfil.uu.se/EUconst.php for EN-PL, EN-GR, EN-DE

Translation models were trained as follows:

English – German
- Europarl
- JRC-acquis
- JRC-Acquis -+ Eu – constitution

English - Romanian
- JRC-Acquis
- SETimes
- Europarl

English – Polish
- JRC-Acquis
- Europarl
- JRC-Acquis + EU Constitution

English – Bulgarian
- JRC-Acquis
- SETimes
- Europarl

English – Greek
- JRC-Acquis
- SETimes
- Europarl
- JRC-Acquis -+ Eu – constitution

2. Cross-lingual retrieval

Work on crosslingual retreival (within Task T 6.4) was concentrated on:

- development of the search engine and adaptation to the ATLAS platform
- automatic extraction of RDF triples from Document annotations provided in WP4.

Planed work for Month 18-24

- Machine translation:
  1. Decision on concatenating or not Europarl and Acquis communautaire and corresponding implications for translation models
  2. Collecton and preprocessing of small domain tageted parallel corpora for all langauge pairs and all domains of ATLAS top-hierarchy
  3. Training of models for all language pairs
  4. Development of domain and POS-factored training models
  5. first integration of the engine with the ATLAS platform
  6. first evaluation round.
- Cross-lingual retrieval
  - finalize the automatic extraction of RDF -triples
  - integration with ATLAS
  - test on English, German, Polish and Bulgarian documents

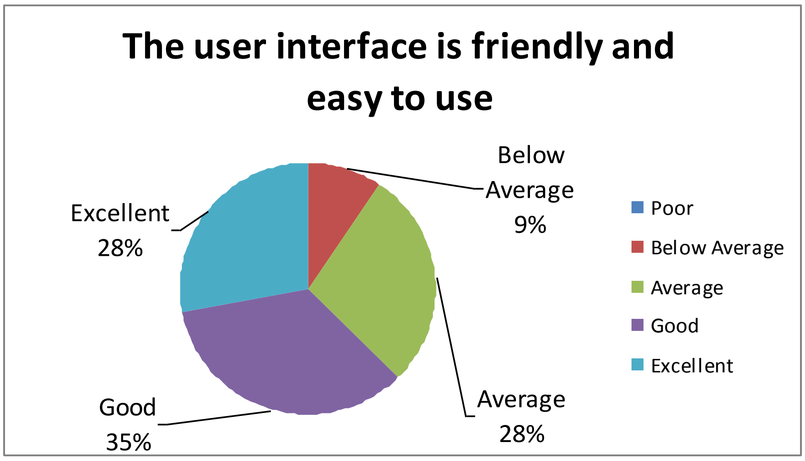
Work package 7 - Testing and User Evaluation

The focus of WP7 is on ensuring the acceptable quality of the final ATLAS system, by assessing the technical performance and the level of fulfillment of the system specifications, as well as assessing the level of user expectations fulfillment. The WP includes 2 tasks; the progress in each one of them during the reporting period is the following:

T7.1: User Evaluation and feedback adjustments

The Task is ongoing. Activities and results within the period:

- The 1st round of the ATLAS system User Acceptance Evaluation has been conducted (following the methodology described in the User Evaluation Plan):
  - 33 users participated from 3 countries (BG=24, GR=7, RO=2).
  - The composition of respondents is: UG1 = 23, UG2 = 6, UG3 = 3, UG4 = 1, UG5 = 0
  - The overall user impression is positive. The MEAN VALUE of each indicator is AVERAGE or ABOVE AVERAGE.
  - For every question in the online questionnaire, a spectrum chart was created, showing the overall picture of user's response per question and reveal any polarisation situations (e.g. in the chart below, 28% of users selected a value of "3" (EXCELLENT), while none selected the value of "0" (POOR)).
  - For every indicator/question "Mean" (μ), "Mode" (τ) and "Median" (m) values were calculated.



- Responding to the project review report, the D7.1 deliverable was updated:
  - The user acceptance methodology was updated to take into consideration "Task Fulfillment Indicators". 21 in total indicators of this type were defined, with the aim to assess user response and performance in certain tasks / steps of every scenario and accompanying exercise. Values for some of these indicators (11) are to be recorded / calculated automatically, through an updated "user logging of activity" system component. This component will record data (e.g. time of execution of a step / scenario / exercise, steps accomplished, etc.) while the user is executing a scenario and an exercise.
  - The Use Case Scenarios were updated to become much more detailed and offer an accompanying exercise. A new detailed scenario (and exercise) was added for the Living Lab users and participating organizations.

  - A new chapter was added to provide more details on the contribution of the Living Labs (LL), as well as on the establishment of the "Multilingual e-Content and e-Library" (MLeCeL) LL.
  - Two LL specific Annexes were added at the end of the deliverable.
- The MLeCeL LL was established in Sofia. As a result of its first activities some new user requirements were identified, such as: "Contact form" widget, "Photogallery grid widget", "Login" widget.

T7.2: System Testing (component, integrated system)

The Task is ongoing. Activities and results within the period:

- Technical Indicators have been drafted for the following system modules. Work is ongoing and all indicators defined so far need to be further detailed in terms of thresholds, pass/fail criteria and testing method (e.g. Test Case).
- CMS indicators (single request / single request in environment of 100 concurrent):
  - CMS_SWDi: time needed to render a static widget i (i = 1..3)
  - CMS_DWDi: time needed to render a dynamic widget i (i = 1..3)
    Need to predefine length / magnitude of associated dynamic content (e.g. a list with 20 entries)
  - CMS_SPi: time needed to render a sample page i (i = 1..3)
  - CMS_SQ: average time and Java heap usage and limits for rendering a search query within a poll of 1.000, 100.000 and 1.000.000 content items.
  - CMS_DB: time needed to render a database response of 100 records.
  - CMS_LG: time needed to respond to a user login request.
- Machine Translation indicators:
  - MT_WDi: % of non translated words for each project language pair (i=1..lang_pairs).
  - MT_BLEU: use widely accepted metrics such as BLEU score, TER, etc and compare against reference values for each project language pair.
- Cross-lingual Content Retrieval indicators (result set of 100 items):
  - CLR_REC: recall rate of the result set (% of docs relevant to query that were successfully returned).
  - CLR_PREC: precision rate (% of retrieved docs relevant to the search).
  - CLR_EXEC: time needed to render a search result of 100 records.
- Summarisation indicators:
  - SUM_RESP: response time to provide a summary of a 10 pages of text in each of the project languages.
  - SUMP_PR: precision / recall rate, based on comparisons between auto generated summaries and gold summaries.

Furthermore, assessment of i-Publisher, i-Librarian and EUDocLib was performed by ATLANTIS, ITD, DCL – IBL and ICS PAS. Detailed feedback and reporting documentation was provided to Tetracom. Finally, ICS PAS gathered user comments from the PALC 2011 Conference; these were passed over to the ATLAS coordinator.

Work package 8 - Dissemination and Exploitation

During the reporting period, the main objectives in Work package 8 “Dissemination and Exploitation” are:

- To promote the ATLAS platform and encourage the use of i-Publisher, i-Librarian and EUDocLib among different target users groups and public.
- To plan project sustainability in order to demonstrate the real value of the ATLAS platform and the online content services, and thus providing a basis for future project funding and extension.
- To disseminate project results through various channels – advertising materials, publications, workshops. Distribution of project leaflets, brochures, conference articles, organization of workshops.
- To present project on both national and international forums.

T 8.1: Project website

The website of the project was built with i-Publisher as an experiment within MLeCeL LL. The list of suggestions for improvements is already in the development list of Tetracom. The website and its vusalisation is already in i-Publisher Simple mode, ready to be used by end users.

T 8.2: Detailed Dissemination Plan

During the reporting period, the dissemination plan was updated including events planned by partners.

1. Tetracom and ITD prepared an application form for the exhibition at the Innovation Convention, 5-6 December 2011 in Brussels (http://ec.europa.eu/research/innovation-union/index_en.cfm?section=ic2011).

2. Tetracom and ITD is preparing to present ATLAS project results at WEBIT’11, 26-27 October 2011 in Sofia, Bulgaria (http://www.webit.bg/). Expected number of attendees is 700 (web developers, e-commerce, hosting, content management and CRM provider, ad network, affiliate system) for networking with colleagues from the region and learn from the best practices and professionals form CEE and the rest of the world.

3. Organisation of workshop “Digital Business Perspectives for the Publishers and Librarians” within European Day of Entrepreneur, 26-27 September 2011 in Sofia. The ATLAS project and i-Publisher, i-Librarian will be presented and demonstrated among representatives from the libraries, universities, publishing houses and wide public. ITD and ATLAS project are co-organizers of the EDE 2011.

4. ICS PAS and Tetracom will present article titled: ATLAS – Multilingual Language Processing Platform by Maciej Ogrodniczuk (ICS PAS) and Diman Karagiozov (Tetracom) at SEPLN 2011: 27th Conference of the Spanish Society for Natural Language Processing in Huelva, Spain, 5-7 September 2011. Project leaflets will be also distributed at the Conference. The main target audience includes IT scientists and linguists. More information: http://www.uhu.es/sepln2011/.

5. Project leaflets will be distributed by ICS PAS and UniZD at the SlaviCorp 2 Conference in Dubrovnik, Croatia, 12-14 September 2011. Organizers of the conference are: Institute of Linguistics, Faculty of Humanities and Social Sciences, University of Zagreb, Croatian Language Technologies Society, META-NET, Multilingual European Technology Alliance, CESAR project (ICT-

PSP), ACCURAT project (FP7) and LetsMT! project (ICT-PSP). Target audience includes IT scientists and linguists. More information: http://hnk.ffzg.hr/slavicorp2011/home.html

6. Organisation of ATLAS Workshop "Integration of multilingual resources and tools in Web applications workshop" @ Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011) in Hamburg, Germany, 26 September 2011. Main organizers of the workshop are Tetracom (Diman Karagiozov), Polish Academy of Sciences (Maciej Ogrodniczuk) and University of Hamburg (Cristina Vertan). The programme committee consists of Damir Cavar (University of Konstanz, Germany), Dan Cristea (University of Iasi, Romania), Svetla Koeva (Bulgarian Academy of Sciences), Vladislav Kubon (Charles University Prague), Lothar Lemnitzer (Academy of Sciences Berlin), Adam Przepiorkowski (Polish Academy of Sciences), Polivios Raxis (ATLANTIS, Ltd, Greece), Roland Winnemöller (University of Hamburg).
   Target audience includes Computational linguists.
   More information: http://www.corpora.uni-hamburg.de/gscl2011/?Workshops:Integration_of_multilingual_resources_and_tools%0Ain_Web_applications
7. Preparation of paper on MT -engine in the ATLAS project, as well as leaflet distribution at RANLP 2011 (11-17 September 2011 Hissar /Bulgaria) and LTC 2011 (25-27 November 2001, Poznan, Bulgaria)
8. Preparation of paper onATLAS project at the main GSCL conference(28-30 September 2011, Hamburg)

T 8.3: Sustainability plan preparation

As the Living Labs approach will be among the main instruments for achieving sustainability of the ATLAS project outcomes, ITD prepared draft exploitation plan to be included in the Sustainability Plan and focused on:

- Defining the initial version of the MLeCeL Living Lab infrastructure and services. The Living Labs approach will be among the main instruments for achieving sustainability of the ATLAS project outcomes;

- ATLAS services pilot based on the MLeCeL LL will be the State University of Library Studies and Information Technologies (SULSIT) and University Computing Centre of Sofia University (UCC);

- MLeCeL Living Lab web collaboration platform is designed to ensure interaction with different ATLAS target users and communities.

- First prototype of the MLeCeL Living Lab web platform is available at: http://livinglab.itd-bg.eu/. The work will continued in the next reporting period;

- Establishing cooperation of MLeCeL Living Lab with other Living Labs.

ATLANTIS provided contribution in the authoring of the draft Sustainability Plan and of the ATLANTIS "Exploitation Strategy and Activities" report.

Task 8.4: Preparation of primary outreach materials

Primary outreach materials includes preparation of brochures, articles for conferences, submission forms for exhibitions and fairs, materials for organization of workshops.

1. Preparation of promotional materials

ITD prepared two brochures for dissemination:

- i-Publisher brochure;
- i-Librarian and EUDocLib cases brochure.

Both brochures are designed to target specific users and to raise awareness among them about the ATLAS services and will be used in the dissemination events organized by ITD.

A new visual presentation of the whole project was initiated by Tetracom. Professional designers and a creative team have started building a new vision of the project and its products and services. The new design will be applied to the dissemination materials, the website, the presentations and involvements in conferences, fairs. etc.

A description of GSCL workshop "Integration of multilingual resources and tools in Web applications" co-organized by Cristina Vertan (University of Hamburg) , Diman Karagiozov (Tetracom) and Maciej Ogrodniczuk (Institute of Computer Science, Polish Academy of Sciences) (http://www.corpora.uni-hamburg.de/gscl2011/?Workshops:Integration_of_multilingual_resources_and_tools%0Ain_Web_applications) was prepared, reviewed and amended.

2. Preparation of conferences papers

   - An article for the SEPLN 2011 conference "ATLAS – Multilingual Language Processing Platform" by Maciej Ogrodniczuk and Diman Karagiozov was prepared and submitted.
   - ATLANTIS submitted a paper abstract to the GSCL 2011 Conference in Hamburg (28-30/09/11): "Towards the integration of monolingual NLP frameworks for multilingual applications: A case study for the Greek Language within the ATLAS project framework".
   - ITD and Tetracom prepared for submission the ATLAS Application form for the exhibition at the Innovation Convention - 5-6 December 2011, Brussels.

Task 8.5: Dissemination of primary materials

The dissemination activities, performed by the partners during the reporting period, include:

- Presentations of the ATLAS project and dissemination of primary materials on national and international conferences, workshops, and meetings.

- Participation in national and international events and exhibitions to demonstrate i-Publisher, i-Librarian, and EUDocLib prototypes among potential users.

- Publications (articles and newsletters) to reach target audiences over a long period of time

- Organization of target workshops to increase visibility of the ATLAS services among potential customers – individuals and organizations and to strengthen the contacts between ATLAS partners and potential members in the ATLAS communities.

- Synergies with other relevant EU projects. Presentation of the Atlas project at project meeting of similar EU projects in order to establish closer collaboration between ATLAS project and other EU projects in the relevant area.

The activities performed by the partners are described in Task 8.8 and Task 8.9 and summarized in Annex 1: Report of Dissemination Activities for the period: 01.03.2011-01.09.2011.

Task 8.6: Interaction with target communities

Atlas project was promoted and i-Librarian, i-Publisher were demonstrated by ICS PAS and DCL – IBL among linguists, computational linguists, NLP engineers, IT scientists , key players in Language Resources and Technologies and representatives of social sciences and humanities.

Atlas project was promoted by ATLANTIS among Business Angels from all over Europe at 11th Congress of the European Business Angels Network (EBAN) in 12-13 May 2011, Warsaw, Poland.

During Usability seminar in May, ITD included i-Librarian and i-Publisher prototypes in the programme for demonstration among potential users - IT experts, PhD students and researchers in Software engineering, web developers. Most of them participated in the first testing round of i-Publisher and i-Librarian first prototypes.

During the reported period, ITD further developed ATLAS collaboration with the following organizations:

- Research and Development Department, Sofia University (http://nis-su.uni-sofia.bg/) is the main organizer of EDE 2011 within which ATLAS will conduct a Workshop.

- Digital Spaces Living Lab (DS LL) (www.digitalspaces.info) and VirtSOI Regional living Lab actively contributed to the development of MLeCeL Living Lab concept and web platform for collaboration.

- During International Workshop: Re-designing Institutional Policies and Practices to Enhance the Quality of Education through Innovative Use of Digital Technologies in June, ITD established closed collaboration with the University of Library Studies and Information Technologies (SULSIT) (http://www.unibit.bg/) for planning future partnership activities with ATLAS project. SULSIT through its UNESCO Interfaculty provided access to its partner organizations:
    - UNESCO Chair, The All Russian State Tax Academy of the Ministry of Finance of the Russian Federation, Moscow,
    - UNESCO Chair Holder, University of Tampere, Finland,
    - UNESCO Chair, St Petersburg State University of Aerospace Instrumentation, Russian Federation
    - The Association of the University Libraries in Bulgaria (AUL)
    - Union of Chitalishta (Municipality Cultural Centres)
    - National Institute of Archeology and Museum, Bulgarian Academy of Science
    - National Library “St. St. Ciryl and Methodius”
    - National Museum of History in Bulgaria

Task 8.8: Nationwide project promotion

ATLAS project was promoted and presented among different users on national level. The nationwide ATLAS promotion activities include:

- Distribution of leaflets at national workshops and seminars, presented in the Annex 1.
- Publications:
  - An article about ATLAS project was published by ICS PAS on the Polish portal for Success Stories: http://en.kpk.gov.pl/index.php?option=com_sobi2&catid=5&Itemid=142&lang=pl. The portal presents Polish successful projects financed under FP5, FP6 , FP7 and the CIP programme.
  - Press release of ATLAS in the electronic newsletter of ATLANTIS (more than 85.000 recipients).
- Demonstration of i-Publisher and i-Librarian – ITD demonstrated ATLAS products at Usability Seminar, organized in 18-19 May in Sofia by Sofia University – 20 participants. The lecturers were Johann Schrammel and Regine Müller from USECON - The Usability Consultants GmbH. The training programme included the following topics: Usability and User Experience, Benefits of Usability, Usability Heuristics and User Interface Principles, Methods of Usability Engineering, Analysis, Innovation, Design, Prototyping, Evaluation, Trends and Special Topics in HCI, Exercises in Usability Reviews & Testing.
- Presentations at National Exhibitions - ATLAS posters in Bulgarian and English are shown at the Permanent exhibition of the Bulgarian Academy of Sciences by DCL – IBL in Sofia.

Task 8.9: Trans-European project promotion

The dissemination activities performed by partners and promoting ATLAS project in a European and international level are also presented in the Annex 1. The activities include:

- Distribution of leaflets and presentation of the ATLAS project at international conferences, seminars and workshops.
- Creation of synergies between ATLAS and other European projects:
  - ATLAS project was presented at CESAR project meeting, 26 June 2011, Budapest, Hungary by ICS PAS and DCL – IBL with the purpose to create synergies between ATLAS project by exchange of ideas and exploring possible ways of partnerships.
  - ATLAS project was presented by ATLANTIS at a Kick-off meeting of the PROMIS Lingua international project in 13-15 April 2011, Luxembourg among business consultants, linguistic, and IT experts, as well as a European SME umbrella organisation.
- Presentations at International exhibitions - ATLAS project and online demo were presented by ICS PAS and DCL – IBL among key players in Language Resources and Technologies at the META-NET

Forum 2011. META-FORUM 2011 – Solutions for Multilingual Europe is an exhibition space in which the participants display and demo various aspects of the work being done across the entire META community and beyond. The exhibition reflects both the research and industry aspects of the community.

Highlight clearly significant results within WP8

The most significant results achieved within WP8 during the reporting period are as follows:

1. Update of Dissemination plan. Dissemination events planned for the next reporting period, including participation in conferences, organization of targeted events, participation in exhibitions and fairs to demonstrate the prototypes of i-Publisher, i-Librarian, and EUDocLib.

2. Performed dissemination activities, summarized in Annex 1.

3. First prototype of the MLeCeL Living Lab web platform is available at: http://livinglab.itd-bg.eu/. Within WP 8 MLeCeL Living Lab web platform will be used for dissemination purposes and as an instrument for ensuring interaction with ATLAS target communities.

Deliverables and milestones tables

N	Name	WP	Lead participant	Nature	Dissemination level	Due delivery date from Annex I	Delivered Yes/No	Actual / Forecast delivery date	Comments
D 1.2	Commercial agreement	1	1	0	CO	M18	yes	M18	resubmission
D 1.3	First year progress monitoring reports	1	1	R	R	M 14	yes	M 14	
D2.3	Software documentation	2	1	O	PU	M15	no	M22	Delay due to additional implementation
D 7.1	User evaluation plan	7	9	O	CO	M18	yes	M 18	resubmission

Table "Deliverables"

Project management

The Consortium established a structure for communication with the EC and the beneficiaries regarding contractual and financial issues, consortium communication, coordination and support of work package leaders, as well as preparation of reports for the EC. In addition, the activities within this package cover regular assessment of project progress and the organization of project meetings.

Deviation from work

- WP 2
  - Implementation of Simple mode
  - The deliverable D 2.3 - Software documentation (due in PM 15) will be delayed due to additional implementation in i-Publisher – Simple Mode. The results after the first user test round clearly showed that a simple mode for the actions in i-Publisher is needed as the system manages complex workflows that should be simplified for a non experienced user. The planned additional work will finish in PM 18 and D 2.3 will be finished in PM 22.
- WP3
  - The Leader of the packaged stepped out of this working package due to the lack of experts in the team and was replaced by Tetracom
- The Croatian NLP is withdrawn. The change affects WP 4, WP 5, WP 6. The planned resources are reduced by PM
- The changes in the resources were done for Tetracom involvement in WP 2, WP 3 and Atlantis in WP 1 and WP 5
- The new table of resources is included in the Amendment and it is attached in Annex 2.

Possible deviation from work ahead

Extension of the number of project-months for ICS PAS in WP5 and WP6: 18 additional person-months for WP5 and 20 additional project-months for WP6 will be used for

Work package 1– Project Management

The tasks included in the first working package, together with their status for the reported period, are listed below.

T1.2 – Commercial agreement preparation

The review report contained a request for resubmission of the commercial agreement. An updated version was resubmitted in PM 18.

T1.3 – Technical communication

Technical communication is an ongoing task.

T1.4 – Project coordination and management

The project management activities took place from M12 to M18. A flow of information is organized as follows:

- management mailing list. The Project CEO and the partners staff executives discuss management topics in the Management mailing list;

- internal reports. An internal reporting system was established and a template for an internal report was approved. It is planned for the internal reports to be issued every three months. However 3rd PM internal report was inapplicable due to the lack of significant activities. The first year report was submitted to the European Commission.
- collaboration area. A project web site internal area is set to enable the partners to work on tasks and topics in protected environment. More details are given in the section Project web site.
- discussion forum. The discussion forum is part of the collaboration area and it is a place for the discussion of technical issues.

T1.5 – Project meetings

One project meeting was held during the reported period. The meeting was in Luxembourg and its focus was the preparation for the yearly review. In addition, the Consortium made an overview of the work done and defined the next steps. The meeting minutes together with all the presentations of the participants can be found on the project web site collaboration area.

T1.6 – Work on contractual deliverables

Resource Monitoring

- The coordinator received the EU grant for the project and distributed the funds among partners. No interest on the pre-financing was yield. Furthermore, the actual state of financial expenditures is constantly monitored through reports, prepared by beneficiaries every three months.
- Quality control and work plan monitoring

The coordinator monitors the project activities, ensuring that they lead to the required deliverables and keep up with the project program. No delays were reported or experienced for the reported period.

- Reporting to the European Commission

The first year progress report was delivered by the Coordinator to the European Commission.

The project planning is kept updated.

Use of resources

Legend: A= actually spent person months for the period 01.03.2011- 31.08.2011, P = Planned for the period 01.03.2011-31.08.2011

	WP1		WP2		WP3		WP4		WP5		WP6		WP7		WP8		Total	
	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P	A	P
Tetracom	8	8	12.90	21.09	1.90	10.14	4.28	4.28	0	0	0	0	1.66	1.66	5.04	5.04	50.23	33.80
DFKI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Atlantis	0.73	0.67	0.34	0.27	0,00	1.80	6.20	6.95	3.78	3.53	0.62	0.72	5.55	5.25	1.19	0.83	18.41	20.01
IBL DCL	1	1	0	0	1	3	6.8	7	1.3	3	4	6	0.5	1	0.5	1	15.1	22
ICS PAS	1.84	1.84	3.00	3.00	5.94	13.37	24.25	24.58	3.97	3.97	1.31	2.75	1.25	1.27	1.95	1.95	43.52	52.73
UHH	0.83	0.14	0	0	0.06	2	1.58	3	0.1	1	16.57	6	0.55	2	2.37	3	30.77	17.14
UAIC	1.9	3	0.15	4	1.56	7	8.61	11	2.18	10	0.58	3	0.78	0.8	1.91	2	17.67	40.8
UniZD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITD	0.1	0	0.5	0	0	0	0	0	0	0	0	0	3.5	2	6.55	8	10.65	10
TOTAL	14.4	14.65	16.89	28.36	10.46	37.31	51.72	56.81	11.33	21.5	23.08	18.47	13.79	13.98	19.51	21.82	161.18	212.9

Table 3.1 Personnel, subcontracting and other major cost items for Beneficiary 1 TETRACOM INTERACTIVE SOLUTIONS for the period 01.03.2011-31.08.2011

Work Package	Item description	Amount	Explanations
	Personnel costs	95 040	
	Subcontracting		
	workshops, dissemination meetings	2 590	
	Bank fee	519	
	Hardware	81	
	Software		
	Server rent	863	
	Deprecations	1 103	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		100 196	

Table 3.2 Personnel, subcontracting and other major cost items for Beneficiary 2 DFKI for the period 01.03.2011-31.08.2011

Work Package	Item description	Amount	Explanations
	Personnel costs	0	
	Subcontracting	0	

	Travel	0	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		0	

Table 3.1 Personnel, subcontracting and other major cost items for Beneficiary 3 - ATLANTIS for the period 01.03.2011-31.08.2011

Work Package	Item description	Amount	Explanations
1, 2, 4, 5, 6, 7, 8	Personnel costs	65.014,91	
	Subcontracting	-	
	Workshops, meetings	1.672,86	• Project meeting + Project review, Luxembourg 7-9/06/2011
WP8	Dissemination events	144,69	• 11th Congress of the European Business Angels Network (Warsaw, May 2011). Attributed only 25% of the total cost
	Deprecations	-	
	Hardware	-	
	Software	-	
	Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT		66.832,46	

Table 3.4 Personnel, subcontracting and other major cost items for Beneficiary 4 IBL DCL for the period 01.03.2011-31.08.2011

Item description	Amount	Explanations
Personnel costs	29505.2	Salaries were paid to the team leader, two programers, four computational linguists, the accountant and for a technical support.
Subcontracting		
Traveling and hardware	1847.41	The amount was spent for the project meeting and some dissemination materials.
Remaining costs	1181.02	Office supplies.
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	32533.63	

Table 3.5 Personnel, subcontracting and other major cost items for Beneficiary 5 – ICS PAS for the period 01.03.2010-31.08.2011

Item description	Amount	Explanations
Personnel costs	91934,71	
Subcontracting	-	
workshops, dissemination meetings	6260,15	
Deprecations	-	

Hardware	-	
Software	-	
Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FINANCIAL STATEMENT	98194,86	

Table 3.1 Personnel, subcontracting and other major cost items for Beneficiary 7 - UAIC for the period (March 2011-August 2011)

Work Package	Item description	Amount	Explanations
1, 2, 3, 4, 6, 7, 8	Personnel costs	5078.96	
	Subcontracting	-	
	workshops, dissemination meetings	1146.10	Project Meetings: Luxembourg
	Deprecations	571.06	A server and a desktop
	Hardware		
	Software	-	
	Server rent	-	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT		6796.12	

Table 3.8 Personnel, subcontracting and other major cost items for Beneficiary 8 University of Zadar for the period 01.03.2011-30.09.2011

Item description	Amount	Explanations
Personnel costs	0	
Subcontracting		
workshops, dissemination meetings	0	
Hardware	0	
Software	-	
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	0	

Table 3.9 Personnel, subcontracting and other major cost items for Beneficiary 9 ITD for the period 01.03.2010-28.02.2011

Item description	Amount	Explanations
Personnel costs	19060.27	Payments to Roumen Nikolov, Krassen Stefanov, Marin Barzakov, Viktoriya Damyanova Ivan Koichev, Stanimira Yordanova , Martin Hristov
Subcontracting		
Travel	781.62	Travel for Ivan Koichev to Luxembourg

Remaining costs		
TOTAL DIRECT COSTS AS CLAIMED IN FIANCIAL STATEMENT	19841.89	

Annex 1: Report of Dissemination Activities for the period: 01.03.2011-01.09.2011

Performed dissemination activities for the period: 01.03.2011-01.09.2011by all project partners within the ATLAS project:

Type of Dissemination Activity	Title	Dates and Place	Participant Name and Institution	Performed Dissemination Actions	Available Dissemination Materials	Target Groups and Number of People Reached
Information portal	Success Stories portal	1 March 2011 Poland	ICS PAS, Poland	Information about the project	Initial information on the project as an article in the portal (http://en.kpk.gov.pl/index.php?option=com_sobi2&catid=5&Itemid=142&lang=pl).	General public *No of people reached: available in the Internet*
Conference	8th International Conference Practical Applications in Language and Computers (PALC 2011)	13-15 April 2011, Łódź, Poland	ICS PAS, Poland	Conference article "i-Publisher, i-Librarian and EUDocLib - Linguistic services for the Web.	Distribution of project leaflets More information: http://palc.ia.uni.lodz.pl ."	Linguists, representatives of social sciences and humanities *No of people reached: 50*
Seminar	Linguistic tasks in ATLAS project (with particular focus on lemmatization of multiword units) @ Natural Language Processing Seminar	18 April 2011 Warsaw, Poland	ICS PAS, Poland	ATLAS presentation, showcase of the multiword prototype	A presentation by Adam Przepiórkowski, Maciej Ogrodniczuk and Łukasz Degórski	Computational linguists, NLP engineers No of people reached: 30
Conference	FLaReNet Forum 2011	26-27 May 2011 Venice, Italy	ICS PAS, Poland	Information about the project	Distribution of project leaflets	Key players in Language Resources and Technologies

						No of people reached: 120
Conference	19th International Conference Intelligent Information Systems and 26th International Conference on Artificial Intelligence	13-14 June 2011 Warsaw, Poland	ICS PAS, Poland	Conference article: Łukasz Degórski: Towards the Lemmatisation of Polish Nominal Syntactic Groups Using a Shallow Grammar on lemmatization of multiword expressions. More information: http://iis.ipipan.waw.pl	Distribution of project leaflets	IT scientists, linguists No of people reached: 100
Project meeting	CESAR project meeting	26 June 2011, Budapest, Hungary	ICS PAS, Poland DCL – IBL, Bulgaria	Information about the project	Distribution of project leaflets	CESAR project members No of people reached:20
Conference	META-NET Forum 2011	27-28 June 2011 Budapest, Hungary	ICS PAS, Poland DCL – IBL, Bulgaria	Information about the project More information: http://www.meta-net.eu/events/meta-forum-2011/exhibition	Presentation of the project leaflets, poster and online demo in the dedicated ATLAS booth.	Key players in Language Resources and Technologies No of people reached: 300
Exhibition	Permanent exhibition of the Bulgarian Academy of Sciences	April 2011 - April 2012 Sofia, Bulgaria	DCL – IBL, Bulgaria	ATLAS posters in Bulgarian and English	ATLAS posters in Bulgarian and English	General public
Workshop	International Workshop: Re-designing Institutional Policies and Practices to Enhance the Quality of Education through Innovative Use of Digital	14-16 June, 2011, Sofia	ITDF, Bulgaria	More information: http://qed.unibit.bg	Distribution of project leaflets	No of people reached: 100 from different EU countries, USA, Russia and Bulgaria

	Technologies					
Seminar	Usability Seminar	18-19 May 2011, Sofia	ITDF, Bulgaria	i-Librarian and i-Publisher prototypes were included in the programme for demonstration among participants	ATLAS project presentation	IT experts, PhD students and researchers in Software engineering, web developers No of people reached: 20
Project meeting	Kick-off meeting of the PROMIS Lingua international project	13-15 April 2011, Luxembourg	Atlantis, Greece	Distribution of EN flyer.	Flyer	Business Consultants, Linguistic and IT experts, as well as a European SME umbrella organisation. *No of people reached: 30*
European Congress	11th Congress of the European Business Angels Network (EBAN). *Theme: "Syndication and co-investment: partnerships and vision for the future"*	12-13 May 2011, Warsaw, Poland	Atlantis, Greece	Discussions on funding possibilities for ATLAS takup. Distribution of EN flyer.	Flyer and brochure	Business Angels from all over Europe. *No of people reached: 100*