SYNC3 Annual Report

 www.sync3.eu

SYNC3 takes media monitoring and tagging to another level by comparing the latest news from traditional media sources and the blogosphere, enabling users to track their evolution and share favourite stories. More specifically, SYNC3 will automatically build a 'thematology' of *news events*, based on a statistical modelling approach, providing brief labels (descriptions) of these events and determining temporal, geographical, and causal relationships between them. Most importantly, SYNC3 will appropriately update the statistical news event models to the blogosphere, allowing the system to automatically find blog posts that comment on these news events, and through a 'sentiment analysis' module determine the blog post author sentiment towards these events. Furthermore, through the use of graphs and maps on an interactive and intuitive user interface, SYNC3 will present a mine of information to the system users, allowing them to update such information individually or collaboratively.

Summary of Activities

The SYNC3 project has now completed its first 21 months. The most important achievements of SYNC3 during this period have to do with the delivery of the first integrated SYNC3 system prototype, including draft versions of the components for news and blogs processing, news event labelling and relation extraction and the User Interface components. These activities supported the successful project demonstration at the ICT 2010 event. On top of that, the project has placed significant effort to produce a manually annotated dataset to be exploited in the ground-truth evaluation of the SYNC3 components. Additional implementation work is planned for the next months in 2011, which will guide the production of a more stable version of the system to strengthen the project dissemination and exploitation activities.

First integrated SYNC3 platform

A first prototype of the SYNC3 integrated platform has been released in 2010, which aims to deliver a platform for aggregating news from various sources, derived from the news world (i.e. news portals, etc.) and the blogosphere, and to provide the end users with sophisticated capabilities with respect to content structuring, management, and delivery. To achieve this, novel approaches have been specified and relevant research work has been conducted, which advances the state-of-the-art in the areas of linguistic processing, textual analysis and data/opinion mining, leading to the release of draft prototype versions of the respective software modules. Apart from the core SYNC3 system, which is a news analysis tool customised to the needs of professional and citizen journalists, SYNC3 embeds the following main components:

- The News Clustering Components
- The Blog Processing Components
- The News Event Labelling and Relation Extraction Components
- The User Interface (UI) Components

News processing Components

The news processing components are exploited to analyse news items from professional news sources and perform the appropriate analysis for categorising these items, based on the news events that are reflected in them. More specifically, the goal is to automatically detect

homogeneous groups of documents that report on the same event by means of clustering techniques. Detecting events allows structuring the news sphere in an effective way and, as a consequence, it should allow the user to access the mass of news articles from an event-based viewpoint rather than a document-based or a website-based one as usually done. Through an innovative model for topic and theme categorisation, news events are efficiently clustered into the existing news taxonomy of IPTC.

Blogs processing and sentiment analysis Components

The project has achieved in delivering a first version of a metadata repository, which manages the annotation of blog posts and their associations with the news events, as they are identified from the news processing components. A blog adaptation model has been developed, which enables the clustering of the unstructured blog posts, according to the structured hierarchy of the news domain. Initial work is also in place for the extraction of sentiments from blog posts, exploiting a hybrid combination of both statistical and rule-based approaches.

News events labelling and relation extraction Components

This set of components exploits the clustering of news items into news events, as it is performed from the news processing components. The aim of the events labelling module is to help users to find out what news events are about. The input into this component is made up of news event clusters containing one or more news items from different sources. The relation extraction module is used to associating news events with each other by identifying their geographical and temporal relations.

SYNC3 User Interface Components

The SYNC3 user interface Components consist of the three main modules:

- The Exploration and Browsing interface module, which allows users to search and explore the underlying collection of news and blogs, as they are classified into news events, in an easy and powerful way;

- The News Story Creator, which allows users to combine news items for creating their own news stories and share them with a group of people;

- The Personalization and Collaboration infrastructure, which enriches the SYNC3 user interface with social features;

The functionalities of the SYNC3 components are well realised through the visualisations provided in the following figures:
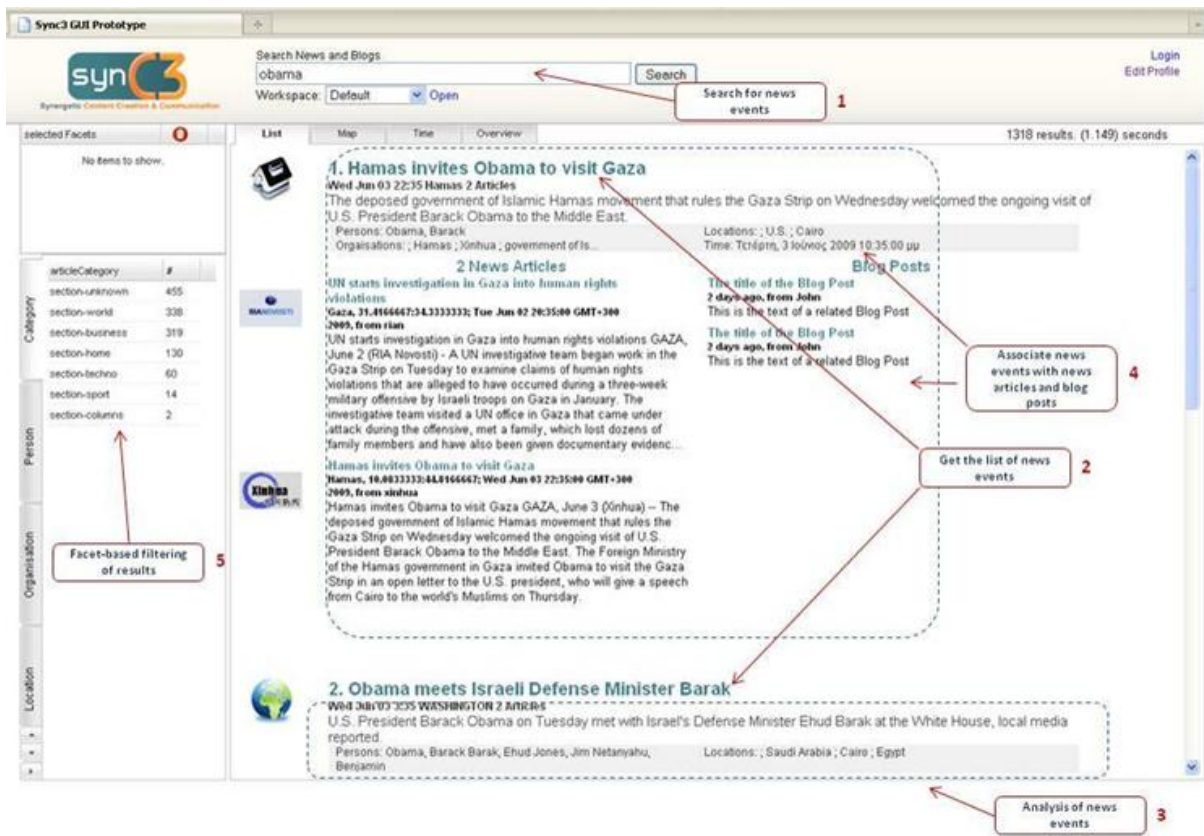
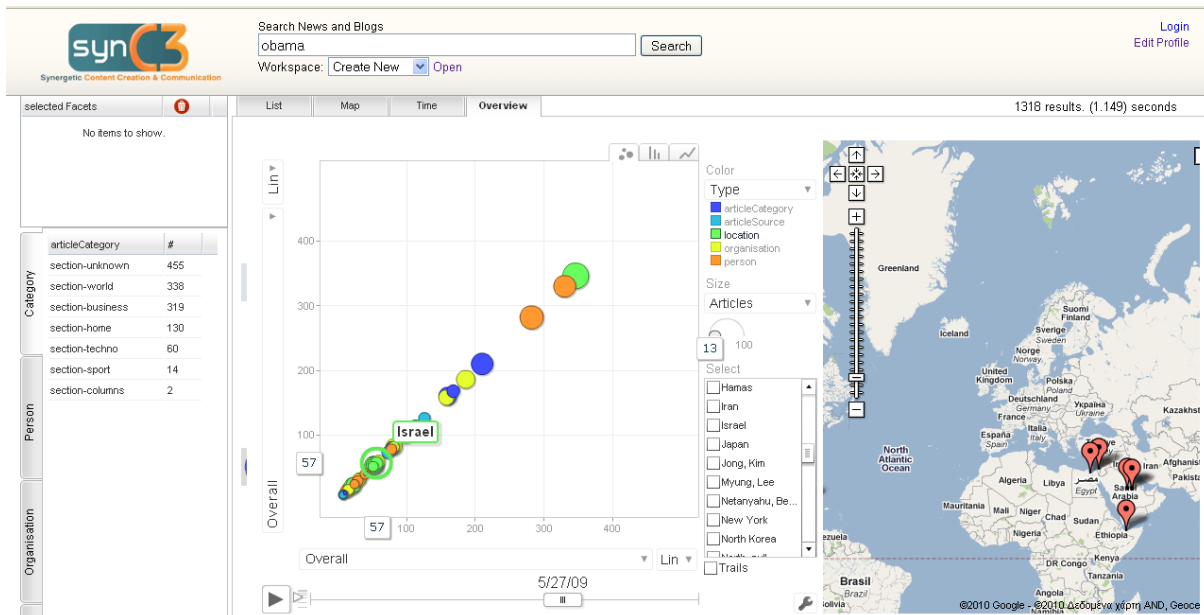**Figure 1: Exploration and browsing of content**

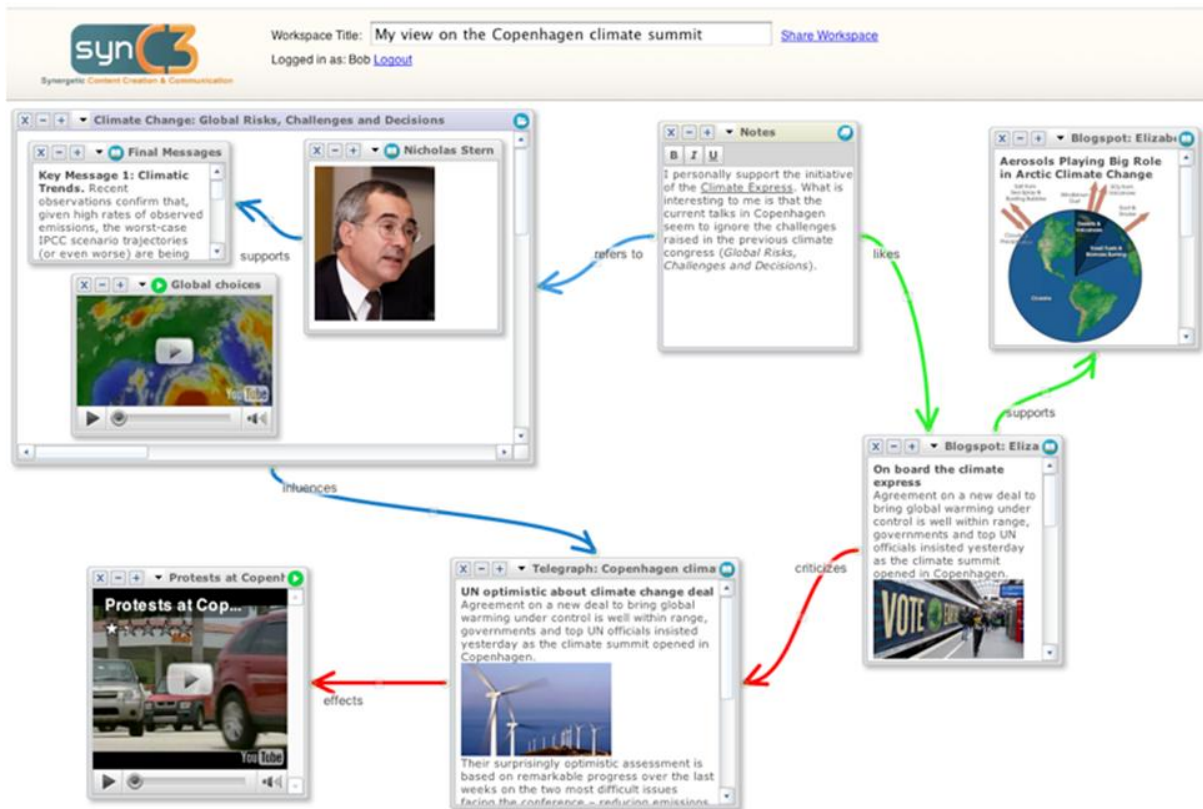**Figure 2: Graphical visualisation of the results**

Figure 3: User-driven creation of stories

Production of annotated dataset

SYNC3 tries to combine a sustainable commercial solution with innovative research in the relevant scientific fields. Thus, the individual components are planned to be tested with different ground truth datasets, so that the results can be comparable to the state-of-the-art in the academic community and enable appropriate dissemination of the research work to the specific scientific communities. On top of that, the project has focused on building a common dataset through the crawling of news and blog sources available on the Internet and carefully selected mainly from the relevant media users of the Consortium, based on multiple criteria. The Consortium has placed significant effort in order to annotate this dataset and exploit it as the main dataset for all SYNC3 modules through the integrated prototype. Finally, collaboration with the LivingKnowledge project has been established to share and reuse this work in the LivingKnowledge project development.

User Involvement, Promotion and Awareness

During this period, SYNC3 has continued with the dissemination and exploitation activities in order to raise the awareness of the target users about the progress and achievements of SYNC3. Through effectively exploiting the main communication channels, such as the project Web Site and the project community in Facebook, LinkedIn, and Twitter, SYNC3 has paid attention to the interested stakeholders to stay tuned on the activities of the project in the last period. Of utmost importance is the fact that a draft version of the SYNC3 integrated prototype has been successfully demonstrated at the ICT 2010 event through the exhibition stand, jointly organised with the EC-funded PAPYRUS and CASAM projects. Furthermore, two SYNC3 presentation videos have been produced: one general project description and one UI presentation video. Finally, based on early drawing of the project competitive landscape, the project is moving towards drafting the project exploitation and business plan.

Future Work

Over the next year, SYNC3 will continue its core implementation work, through a rapid prototyping approach. By doing so, the strong target users' engagement is planned, which will enable the actual validation of the first prototype version and the SYNC3 concepts. This work will guide the delivery of a second one, incorporating updates on the individual components and the platform as a whole through this evaluation phase. In the same year, the important and time and effort consuming task of manually annotation of the SYNC3 dataset is foreseen to be completed, which will offer flexibility over the ground-truth evaluation of the components. In addition to the further developments, the project will continue approaching interested users for disseminating its concepts and results, and for exploring business opportunities to allow SYNC3 penetration into targeted markets and user groups.

Further Information

- **www.sync3.eu**
- **SYNC3 Press Release**
- **http://cordis.europa.eu/fp7/ict/content-knowledge/sync3_en.html**
- **http://twitter.com/sync3**