2. Publishable Summary

2.1. Project Objectives

SYNC3 aims to apply the news domain structure derived from well-organised news portals to the unstructured domain of the blogosphere. To achieve this, novel approaches are proposed in each of the research areas of the project, advancing the state-of-the-art in each area and producing a number of software modules that will be integrated into a common platform operating as a news aggregation tool. This tool will be organizing content coming from both news portals and blogs. It will also allow the creation of more user-generated content, either by authoring new material, or by re-organising the links structured by SYNC3 into user-generated storylines.

The objectives of the SYNC3 project are the following:

- Identify correlated text excerpts among news articles to provide the news thematology, by grouping news items into events, topics, and themes.
- Appropriately adapt the news event models and use them to find blogs that comment on the corresponding news events, while performing sentiment analysis on them.
- Use existing tagging and annotation information to extract news event labels, as well as geographical, temporal, and causal relations between news events.
- Visualise a graph and its rich information in a straightforward and easy to navigate way, providing users with the functionality needed to, individually or collaboratively, augment this graph with content and viewpoints.

2.2. Expected Results

SYNC3 has three main goals and target groups. First, it helps journalists in efficiently exploiting the unstructured blogosphere for their everyday work. Second, it gives bloggers a new platform to share news and opinions with other content creators. Third, it helps communication professionals and policy makers to better follow the public debate.

SYNC3 aims to deliver a user-friendly news analysis tool for searching blogs and traditional media news, allowing users to create, comment and "sync" their news in a virtually limitless network. This tool will integrate functionalities laid on three areas, namely news clustering, blog processing and news events labelling and relation extraction, and will be customisable to the needs of the professional and citizen journalists.

More specifically, the project will deliver the following:

- The SYNC3 system, which is going to be released as a news analysis tool customised to the needs of professional and citizen journalists.
- The news clustering components, providing modules for news article clustering and their classification into events, as well as algorithms for dynamic and hierarchical clustering.
- The blog processing components, delivering modules for blog post classification, models for sentiment analysis, and methods for knowledge transfer from the news to the blogosphere domain.
- The news event labelling and relation extraction components, which include modules for cluster labelling and relation extraction and methods for news relations and attributes extraction.

- The user interface (UI) components, providing the appropriate visualisation techniques and support for user interaction and personalization.

2.3. Work Performed and Results Achieved

This section provides an overview of the achievements during this reporting period of the project, grouping them into technical, dissemination, exploitation, and management results.

Technical Achievements:

- Release of the final version of the topic and theme categoriser.
- Release of the final version of the event recognition module.
- Evaluation of the WP3 chain components using the SYNC3 corpus and report on scalability issues.
- Release of the final version of the querying and reasoning component.
- Release of the final version of the blog adaptation and classification module.
- Integration of the sentiment analysis tool into the SYNC3 system.
- Release of the final version of the sentiment analysis tool.
- Evaluation of the WP4 chain components using the SYNC3 corpus and report on scalability issues.
- Improvements on the performance of the news event labelling and relation extraction processing modules.
- Release of the final versions of the news event labelling and relation extraction processing modules.
- Visualisation of stories with graphs based on similar events and timelines showing key players.
- Evaluation of the WP5 chain components using the SYNC3 corpus and report on scalability issues.
- Release of a redesigned version for the News Story Creator.
- Updates on the exploration and browsing and the profiling infrastructure components, based on the user evaluation results.
- Development of an iPad SYNC3 application for the exploration and browsing functionality.
- Release of the final integrated prototype for user testing.
- Organisation of the intermediate and the final user evaluation sessions for assessing the perception of the target users on the suitability and quality of the SYNC3 platform.
- Report on the final integrated prototype and the scalability issues.

Dissemination Activities:

- Maintenance of the project Website.
- Preparation of a joint paper and presentation of SYNC3 to the eChallenges 2011 conference.
- Continuous dissemination activities, through publications, presentations and participation in relevant events.
- Maintenance of various dissemination and communication channels.

- Communication of project achievements to relevant research communities and commercial stakeholders.

Exploitation Activities:

- Submission of the exploitation and business plan.
- Preparation of the SYNC3 IPR Agreement.
- Presentation of project results to target commercial stakeholders.
- Submission of patent applications.

Management Activities:

- Continuous project management activities.
- Finalisation of project activities.
- Preparation for the final review.

2.4. Impact

SYNC3 aims to deliver a user-friendly news analysis tool for searching blogs and traditional media news, allowing users to create, comment, and "sync" their news in a virtually limitless network. SYNC3 integrates functionalities in three areas, namely news clustering, blog processing, as well as news event labelling and relation extraction, in a fashion customisable to the needs of the professional and citizen journalists. SYNC3 is developing novel and fine-tuned methods to connect users to news content and other users. Providing this kind of social activity enabling functionality may prove a turning point for news media and social networks.

Countless attempts have been made to provide centralized repositories of personal content. However, blog users do want to have control of their own content and are often unwilling to get locked-in to a specific provider. This leads the blogosphere to be a decentralised and largely unstructured universe. SYNC3 aims to tackle this problem by structuring the part of blogosphere that refers to running news stories, rendering it accessible, manageable, and re-usable.

Notably, SYNC3 is a step towards a paradigm that represents a radical departure from news reporting as we know it today. By observing the facts and trends in internet news media and social networks, we arrive at the conclusion that today's environment is not only ready to accept the change, but also demands it. All the pieces of the puzzle are already there. The one thing missing is a technological platform that will provide the required synergy, by addressing the long-standing issues of user-created content management and presentation.

SYNC3 aims to render the vast blog content referring to current news easily accessible, thus lifting the last barrier in the way towards a new era of effective communication among citizens and collaborative formation of public opinion, an idea that has long been evangelized by media and internet experts alike. SYNC3, as an enabling platform, is the last piece in the puzzle needed for materializing the concept of collaborative public opinion formation.

2.4.1.1 Scientific Impact

SYNC3 does not depend on the blog content to derive the needed information structure, but instead employs the more structured news articles to derive a news thematology. This general concept of deriving the thematology from one content and then applying it to another, less structured content is an innovative approach in SYNC3. The SYNC3 news article clustering methodology is particularly suited to the news domain. It defines news events as the basic abstract common interest "atoms". The method does not categorize the news content in a specific pre-defined ontology and neither does it extract the underlying semantics. Yet, the linguistic statistical clustering of SYNC3 generates clusters that are precisely specified and

logically separated. That is, the combination of the statistical approach and the specific regularities in news articles distils the human knowledge that is embodied in the (statistical) way the news articles are structured.

The adaptation of the news article clustering results to a new domain (that of the blogosphere) constitutes a major innovative research direction in SYNC3. The linguistic and structural differences between news articles and blog posts are neither large enough to render the results of news article clustering totally inapplicable to blog posts, nor small enough to allow the aforementioned results achieve the desired level of performance. The differences between news articles and blog posts are systematic in nature and therefore can be captured in a method that augments or updates the news article clustering results, in order for the latter to achieve the desired performance in the blogosphere domain.

Research on extraction of relations from text has focused in identifying relations between named entities. In the context of SYNC3, the relation extraction methods adopted identify relations between excerpts in text. Each excerpt usually has multiple named entities, and many of them are also present in both excerpts. The pioneering work in this research area addresses the fuzziness and inaccuracy in its input, i.e. in the determination of the excerpts. The method aims to compensate these drawbacks by working with the multiple available articles for each relation and subsequently augmenting the results to phase out the errors introduced by inconsistencies in individual articles.

Sentiment analysis faces similar challenges. Traditional sentiment analysis research primarily focuses in detecting opinion about named entities. The method in the context of SYNC3 operates instead on excerpts that have multiple named entities. To enable this, SYNC3 work adopts various strategies, such as holistic approaches that group excerpts across each polarity, instead of analyzing each excerpt independently.

The human-computer interface breaks new ground to achieve the visualization of the information-rich graph. There exist multiple types of nodes and connectors in this graph: news events, topics, themes, users, user content, relations (causal, temporal, geographical) between news events, opinions per user attached to news events and possibly grouped to opinion groups, existing connections between users based on interest, etc.

2.4.1.2 Socioeconomic Impact

SYNC3 delivers a platform that can be commercially exploited to analyse existing news and blog sources and provide a comprehensive roadmap to news creation, relevant opinions, and sentiments expressed for news events in the blogosphere. Through SYNC3, media organisations and other relevant stakeholders will be able to monitor the social media environment and participate in fostering local, national, and international growth. SYNC3 will offer the capability to monitor the news sources and correlate events with non-professional opinions. The available information sources, including news agencies and media organisation portals, as well as the macro- and micro-blogging sphere, will be exploited to discover the trends with respect to the news events covered by the social media. The corresponding SYNC3 stakeholders could also engage target markets in the media sphere by becoming content “prosumers”, thus providing a completely new area for economical transactions. SYNC3, as a collaborative platform, could assist in measuring the impact of news and enable media organisations to refine their strategy. It will be used as a platform offering a totally new experience of news creation by providing the environment to set criteria in delivering real news events to the wide public, as well as to listen to the public feedback about the events occurring around them, in order to reform the way that media content is produced and distributed.

2.5. List of Beneficiaries

no.	Beneficiary Name	Beneficiary Legal Name	Short Name
1	Athens Technology Center	Athens Technology Center S.A.	ATC
2	National Centre for Scientific Research "Demokritos"	National Centre of Scientific Research "Demokritos"	NCSR'D'
3	Leibniz Universität Hannover	Gottfried Wilhelm Leibniz Universität Hannover	L3S
4	The University of Edinburgh	The University of Edinburgh	UEDIN
5	Xerox Research Centre Europe	Xerox SAS	XRCE
6	European Journalism Centre	Stichting European Journalism Centre	EJC
7	i-Sieve	I-Sieve Technologies Ltd	ISIEVE
8	RIA Novosti	Russian News & Information Agency 'RIA Novosti'	RIAN
9	Google	Google Ireland Limited	GOOGLE

2.6. Coordinator Contact Details

Company Name: **Athens Technology Center S.A**

Address: **10, Rizariou Str. 152 33, Halandri, Athens, Greece**

Tel: **+30 210 6874300**

Project Coordinator: **Nikos Sarris**

E-Mail: **n.sarris@atc.gr**

2.7. Project Logo

2.8. Project Website

The project public Website is accessible through: www.sync3.eu