WikiTimes is a project in the context of the L3S Web Observatory show case that aims at building a structured and rich knowledge base of news events by harvesting the efforts of the Wikipedia crowd. Adequate interfaces will be developed for accessing the structured news events and for the integration with existing knowledge bases such as DBpedia and YAGO.

This research project aims at bringing structure to news events in Wikipedia in order to help users find answers to “What, Who, When, Where, and Why” questions about news topics. Summaries of news events from the past 13 years will be extracted from Wikipedia and indexed in a knowledge base, which will be then exposed to the end users via an intuitive search interface.

While traditional search engines can effectively deliver a load of news articles related to a query, they fail in finding objective information, removing duplicates and providing the users with a timeline summary of important news stories. Since more than 13 years, the Wikipedia community has started a sort of crowdsourcing-based effort for manual summarization of news events in the so-called Current Events portal of Wikipedia (http://en.wikipedia.org/wiki/Portal:Current_events). Nevertheless, the lack for a well defined structured access to the Current Events wiki pages (as an inherent problem of Wikipedia), limits the extent to which this very valuable content is exploited.

In order to help users find answers to “What, Who, When, Where, and Why” questions about news, WikiTimes aims at extracting and indexing various relation types for enriching the events data such as temporal, causal, spatial, topic hierarchical relations between events as well as entity-event relations.

In particular, an ontological framework for representing news events will be developed and implemented to build a rich news events knowledge base. The knowledge base will be then exposed via a SPARQL endpoint for processing queries about events and entities. Furthermore, a tool for producing timeline summaries for news stories as well as histories of entities involved in the news stories will be developed.

In addition, the summaries of news events that are extracted from Wikipedia’s Current Events portal will be enriched with additional web resources such as related news articles, multimedia resources and posts from micro-logging platforms such Twitter in order to add more context and capture more (diverse) aspects of the events.

The target users of WikiTimes service are both average persons, who are looking for temporal summaries of relevant news stories, as well as researchers, who are interested in analyzing and exploiting the knowledge base of WikiTimes for improving and evaluating temporal information retrieval and summarization techniques.
http://wikitimes.l3s.de/

Research Area:
Intelligent Access to Information

Status of the Project: