

Situational Web Synthesis

Julien Tscherrig

Nowadays, the search for information on the Internet is done through search engines. Search engines return information, present on the Web, indexed based on one or more keywords entered by the user. The results of these search engines are then available in the form of a list of links to a web page that is supposed to contain the information one is looking for. This thesis proposes an alternative approach that provides the information sought directly to users in the form of a synthesis. This synthesis is automatically built based on the results provided by existing search engines. To allow the construction of this synthesis three main themes have been addressed.

The first theme concerns the aggregation of documents on the Internet for a given context or search. Based on this collection of documents, an extractive summary allows the user to visualize the important information for this context or search topic. To build the extractive summary, it is necessary to semanticize the data present on each page before merging them. The data semanticization is carried out either in a parameterized way (e.g. for a given format type) or automatically (e.g. for plain text). The merging process consists of determining the distance between two data graphs and deciding whether or not to merge them. To achieve this synthesis, two sub-themes are studied: the extraction of unstructured data from the Internet and the recording of this data in a homogeneous structured format and the determination of the quality of a document on the Web using objective and measurable criteria.

Several concepts and prototypes have been created to validate this thesis work. The main field covered revolves around Swiss politics and many resulting concepts have been validated and used for several years by the Documentation Service of the Federal Assembly (Swiss Parliament). These concepts made it possible to propose an analysis of the information published on social networks during federal referendums. These concepts have yielded positive results and have validated the approach proposed in this thesis work.

Jury:

Prof. Dr. Philippe Cudré-Mauroux (thesis supervisor)

Prof. Dr. Omar Abou Khaled (thesis co-supervisor)

Prof. Dr. Elena Mugellini (external co-examiner)

Prof. Dr. Hatem Ghorbel (external co-examiner)

Prof. Dr. Rolf Ingold (internal co-examiner)

Prof. Dr. Denis Lalanne (president of the jury)