

access to the Web of Data that can be used by developers and businesses to power products and services. Interestingly, the Sindice crawler is not limited to crawling RDF-graphs but also crawls other forms of structured data, such as those using the HTML5 microdata format using the schema.org vocabulary, or other microformats.

Sig.ma is a user interface that was created as a demonstration of live querying on the Web of Data. Starting from keyword search, Sig.ma identifies the URLs on the Web of Data that are relevant to a keyword and finds all triples that describe these URLs. Through an interactive user interface the user can investigate which triples originate from which data sources, select or remove specific data sources, “walk the graph” by clicking on the URLs which are given as related to the search term, and turn on the URLs related to those, etc.

By using the Sindice search engine, Sig.ma gives the user a live view of what is on the “Web of Data” about any given topic. Due to its ingenious user interface and to the very complete database of Sindice, Sig.ma is probably the best location for lay users to get a feel for the structure and size of the current Web of Data and the sheer amount of knowledge it already captures.

6.7 OpenCalais

The press agency Thompson Reuters produces thousands of news items every day, covering wide areas of science, business, politics, and sports. Although originally intended to be read by humans, a key observation of Reuters is that many of the consumers of their news items are no longer people, but instead computers. Computers are reading the news flowing out of Reuters, analyzing it and producing summaries, reading lists, financial investment advice, etc.

This leads to the paradoxical situation of computers producing information for computers using unstructured natural language as a rather unsuited communication medium. Reuter’s OpenCalais service is intended to help computers process data cap-