Abstract

The occurrence of a noticeable number of cancer incidence leads to requests for cancer cluster investigations. Epidemiologists have to investigate if these number of cancer incidence are higher than the expected value and not coincidental. External factors like certain environmental substances play a role in the development of cancer and have to be considered during the investigations.

Such environmental information is available online but often distributed and of heterogeneous format. Current epidemiological cancer cluster investigation approaches are often unstructured web searches of potential emission sources, this work examines how linked spatio-temporal data can help improve this situation. In order to answer this question the thesis presents a prototypical tool that supports specialists and non-specialists of semantic information systems to make a first survey of cancer-suspicious regions with focus on environmental data. The developed prototype application serves as exploration and information tool for the specific epidemiological use case. For this, the work focuses on cancer cause-effect related Linked Data derived from distributed, isolated environmental and epidemiological data. Another outcome is the domain ontology that reveals the cause-effect chain that is based on International Agency for Research on Cancer (IARC) knowledge: cancer type (e.g. lung cancer) - carcinogen (e.g. diesel exhaust particulates) - emission process (e.g. driving) - transport way (e.g. aerosol) - emission source (e.g. car) - exponent (e.g. males/females).

The concept and first functional requirements of the prototype were defined in collaboration with the Institute of Epidemiology and Social Medicine (IES) Münster. The application development process was conducted in an agile, iterative manner. Feedback of meetings were integrated directly into the development process. Semi-structured interviews of participants with epidemiological background and expert interviews with experienced people in the field of GIS and semantic technologies evaluated qualitatively the final version of the Linked Data based prototype.

One key result is that the previous unstructured recherche of cause-effect knowledge from IARC monographs is structured, linked with spatial information and can be accessed by applications. Another key results is that the presented prototype can support spatio-temporal data exploratory through Linked Data querying and background information recherche tasks by discovering related links of resources in cancer cluster investigations. Depending on the linked data sets involved and the subject of the cancer cluster requests. The application was named as rich by experts and intuitive predominantly by epidemiological users. Experts and epidemiologist stated that the use of the application could be learned in a manageable amount of time.