Leadsto - Collaboratively Constructing and Discovering Causal Chains

Mark Hefke
Forschungszentrum Informatik (FZI)
Haid-und-Neu-Straße 10-14
76131 Karlsruhe, Germany
hefke@fzi.de

Andreas Abecker
Forschungszentrum Informatik (FZI)
Haid-und-Neu-Straße 10-14
76131 Karlsruhe, Germany
abecker@fzi.de

ABSTRACT
Leadsto is a prototypical Semantic Portal for collaboratively describing statements of the form "x leads to y" (e.g. "accident leads to traffic jam"). Existing elements of statements (precedents, antecedents) can be linked with each other, and completely new elements can be created. Individual statements can be created and the set of stored statements further extended and developed collaboratively on the Web by humans; in addition, automated approaches for extracting further statements from any web page are employed. The constantly growing net-like structure can be searched and navigated. The major benefit of the system is to automatically discover and make available causal chains of the form "x leads to y "," y leads to z ", etc. (as well as the reverse direction). In this way, not yet known facts as well as their provenance can be collaboratively discovered by the wisdom of a crowd.

Categories and Subject Descriptors
H.4 [**Information Systems**]: Information Systems Applications

Keywords
Causal Chains, Lightweight Semantics, Social Software

1. INTRODUCTION
No person has the ability to completely overlook complex cause-effect relationships. Quite often, sequences of unfortunate events or misinterpreted signs cause the occurrence of unexpected incidences. This phenomenon can, e.g., be observed when examining air accidents or quite simply when trying to get an overlook about complex procedures, where different people have different and above all limited views on a specific issue. Otherwise, if one only would have a complete overview on a given situation, complex issues could be much easier to understand. Thus, the basic idea of this demonstrator is to provide an easy-to-use tool for collaboratively developing and structuring statement chains, such that a number of people can develop complex interrelations (causal chains) by only providing small pieces of information.

2. SOLUTION APPROACH
Due to the fact that semantically enabled tools usually deliver a significant difference in quality not until a specific amount of knowledge is formalized, we started with auto¬matically generating an initial knowledge base by using traditional information retrieval tech¬niques. In this context, we at first parsed search results of directed queries to web search engines (e.g., by simply using the search string "leads to" or synonyms) and freely available encyclopedia in order to automatically extract state¬ments of the type "x leads to y". In a next step, we extended the existing knowledge base by identifying causes or effects for existing elements in order to retrieve further causal chains, Furthermore, we used translation services to translate facts between the lan¬guages German and English as well as thesauri, in order to identify synonyms and hierarchical relations. Finally selected users were asked to enhance or correct the initially generated knowledge base by collaboratively using the de¬ve¬loped demonstrator. Technical components and functionalities of the system are sketched in the following section.

3. TECHNICAL COMPONENTS
Our ontology-based software architecture comprises the following components:

1. An ontology-based **Knowledge Base** structures statements and interrelations between them.
2. A **Statement Construction Component** supports the user in constructing new statements which consist of element pairs.
3. **The Causal Chain Identification** identifies causal chains and cycles in the knowledge base which are not explicitly modelled.
4. **Administration Functions** support the maintenance of the knowledge base.

3.1 Ontology-based Knowledge Base
Each statement is stored as a set of two interlinked "element instances" in an RDF-based KAON Ontology Instance (OI) Model[1]. The conceptual level of the ontology contains the concept *Element*, the property *leadsto* and its inverse property *is attributed* (cf. Figure 1). These two properties can be used to construct statement pairs. Similar to the SKOS Core Vocabulary [1], the property *broader*, as well as its inverse property *narrower,* can be used for expressing that an element is more generic or specific than another one.

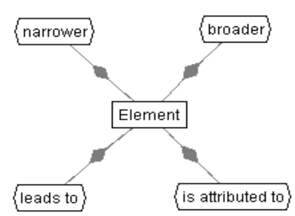

Figure 1. Ontology Scheme

[1] Cf. http://kaon.semanticweb.org/

3.2 Statement Construction Component

Using the construction component, users can collaboratively construct new statements out of existing elements, by newly describing and connecting two elements, or even by combining both options. Moreover, hierarchical relations between elements can be created. When storing new statements into the ontology, already existing elements are automatically discovered, i.e., only new elements of a statement are stored into the ontology and linked to existing elements using the property *leadsto*.

3.3 Causal Chain Identification

The statement identification component checks for every single element the availability of direct successors (and, transitively, for successors of identified successors). Finally, it returns a causal chain in the form of a tree diagram (cf. Figure 2). E.g. the component identifies, that *Cell Phone Radiation* leads to a *sleep deficit*, which results in *overweight*, which in turn leads to several other *diseases* (e.g. *gonarthrosis*). Moreover, the component considers instantiations of the property *broader* in order to infer further possible statements. If, e.g., the element instance *cell phone radiation* has a broader instance *radiation*, then all consequences of the element *radiation* (e.g. *cancer*) are also presented as a potential consequence of *cell phone radiation*. This leads in this example to the fact that the component also identifies *cancer* as a consequence for *cell phone radiation*. The causal chain identification can also be processed by regarding the reverse direction. Based on that, the system can be used for discovering causes of certain circumstances.

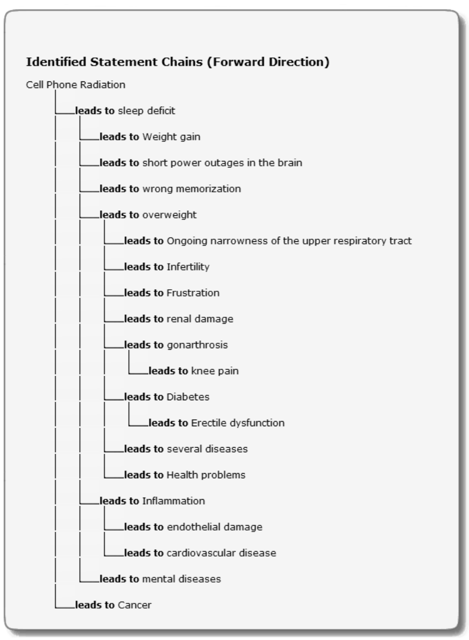

Figure 2. Identified Causal Chain

3.4 Administrative Functions

Administrative functions provide means for automatically learning new statements by parsing search results of arbitrary web search engines. Similar or equal elements can be automatically discovered and merged with simultaneously considering interrelations to other elements. In order to retain background information for single elements, (Wikipedia) Web Links can automatically be discovered. Elements can automatically be translated into different languages. Synonyms and hierarchical relations between elements can be discovered based on dictionaries like WordNet[2].

3.5 Related Work

We combined approaches from Lightweight Ontology Engineering and Information Extraction with the principle of causality. In literature, there exist different tools for web-based collaborative engineering or development of knowledge bases (cf. [2],[3]). From a technical point of view, the identification of causal chains, namely the small reasoning component of the prototype does not need to anything which a standard reasoner could not also do. Causality influences various fields like philosophy, science, physics, engineering or medicine for hundreds of years [4],[5]. Furthermore, a couple of methods for visualizing causal chains (e.g. fishbone diagrams, cause-and-effect diagrams, etc.) do exist. However, there currently exists no easy-to-use tool for collaboratively constructing and discovering cause-and-effect chains.

4. CONCLUSION AND FUTURE WORK

We have described a system which technically supports the collaborative construction of causal chains. A first system prototype disposes of 3300 elements (2800 synonyms) and 2700 statements. In addition, more than 900 hierarchical relations between elements do exist, that are used for inferring further causal chain links. The demonstrator can be assessed on:

http://leadsto.fzi.de

For the future, we intend to validate the developed demonstrator in a specific domain (e.g., medicine or economics). In addition, means for collaboratively evaluating relations between elements are planned. In order to automatically identify more reasonable causal chains, the statement identification component will be extended by an automatic disambiguation of elements based on their position in the hierarchy and the consideration of user-evaluated relations.

5. REFERENCES

[1] Brickley and Miles: SKOS Core Vocabulary Specification. W3C working draft, W3C, November 2005.

[2] Zacharias, V., Braun, S.: SOBOLEO - Social Bookmarking and Lightweight Ontology Engineering, Workshop on Social and Collaborative Construction of Structured Knowledge (CKC), 16th International World Wide Web Conference (WWW 2007), Banff, Alberata, Canada, May 8-12 2007.

[3] Krötzsch, M., Vrandecic, D. and Völkel, M.: Semantic MediaWiki. In I. F. Cruz, S. Decker, D. Allemang,, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, ISWC, volume 4273 of Lecture Notes in Computer Science, pages 935–942, Springer, 2006.

[4] Mackie, John L.: The Cement of the Universe - A Study of Causation. Oxford: Clarendon Press, 1980.

[5] Salmon, Wesley C.: Causality and Explanation, New York/Oxford, 1998.

[2] Cf. http://wordnet.princeton.edu/