6M-9

Meta-Structuring of Concept Chemical Representation (CCR), a Natural-Language-like Knowledge Representation

GIAN Yi-Hsen Hironori Tomobe Hitoshi Iba Mitsuru Ishizuka
Dept of Info & Comm Eng, Univ of Tokyo

1 Introduction

Concept Chemical Representation (CCR) provides us with a means to express knowledge using words as a building block linked together with various specific relations to form a single knowledge or fact. This representation allows us to gather a group of facts to use as a knowledge base to perform various inferences. However, as with any significant amount of knowledge in the real world, it is neccessary to organize the knowledge in a reasonable manner to enable easier comprehension as well as increased focus of the knowledge. In addition, it is much more efficient and relevant to concentrate on a specific domain of knowledge than to gather all the available knowledge altogether to perform any form of inference. Knowledge sharing, which is one of the main proposed uses of CCR, implies that more than one party could be providing the knowledge used. Merging all the knowledge into a huge pool would cause confusion, and make it extremely difficult to maintain the knowledge base. Furthurmore, the separation would allow users of the system to view the difference or similarities of various groups of knowledge.

Thus the need for a Meta-Structure when using CCR. This Meta-Structure helps organizes the available knowledge into separate 'pages' and allows inference to be performed over multiple 'pages'.

2 Basic CCR

Here I would briefly introduce CCR which is the knowledge representation used in this system.

2.1 Atomic & Molecular Concepts

The building block of this knowledge representation is the *atomic concept* which is single words in natural language that represent simple concepts. An object like 'apple' or an action like 'eat' are examples of *atomic concepts.*

Atomic concepts only represent simple ideas or objects and are inadequate to represent real-world ideas or objects, thus there is a need to express more complex ideas using other atomic concepts. *Molecular concepts* are multiple atomic concepts combined to represent a complex idea. In a *molecular concept*, there is a main atomic concept which is 'modified' using other atomic concepts. *Labels* define the manner in which the main concept is 'modified'.

$$(girl\ (MOD\ beautiful))$$

is an example of a molecular concept where the *label* MOD is used with the atomic concept *beautiful* to 'modify' the main atomic concept *girl* to represent the idea of a "beautiful girl".

2.2 Concept Relation Formula

Concept Relation Formula (CRF) represent complete ideas or facts by relating two molecular concepts with a *concept relation.* These *concept relations* will determine the way the two molecular relation interact with each other. Each *concept realation formula* will generally correspond to a sentence in natural language.

$$((Peter)\ [DO]\ (eat\ (OBJ\ apple)))$$

represents the knowledge "Peter eats an apple". *Concept relations* are ordered relations where the molecular concept to the left and right of the relation cannot be reversed without changing the *concept relation* correspondingly.

2.3 Inference with CCR

In CCR, when a query is made against the knowledge base, it will not only match knowledge that is explicitly expressed. An attempt is made to search for knowledge that might satisfy the query through the merger of two different CRFs. This merger is only possible when there is a common molecular concept as well a *concept product* from their respective concept relation.

$$((Sparrow)\ [BG]\ (bird)) + ((bird)\ [DO]\ (fly))$$

$$\mapsto ((Sparrow)[DO](fly))$$

In the above example, the inference is makes use of the common molecular concept "bird" and the *concept product* of $[BG] \times [DO] \mapsto [DO]$.

3 Meta-Structure

The main objective of the *meta-structure* is to provide a means to organize or group knowledge into managable chunks. However, it is also neccessary to be able to maintain the ability to include knowledge from different domains to increase the breadth of the knowledge used for inference. The focus of this *meta-structure* is to provide the means to store the knowledge in segregated groups and to perform inferences that involve other domains as neccesary.

3.1 Pages of Knowledge

Each chunk of knowledge is placed in a *page.* In this *page*, other information like the natural language source of the knowledge and the author of the knowledge. Although these information may not be strictly necessary for inference, it is useful when presented to

[0]概念化学知識表現 (CCR) のメタ構造化機能
ギャンイーセン、友部 博教、伊庭 斉志、石塚 満
東京大学工学部電子情報工学科

the user, facilitating easier understanding of the result derived. In addtion to these auxilary information, there is also links in each page to create a network of *pages* that are used to expand the knowledge base for inference when neccessary. The following figure 1 graphically describes the structure of each *page*.

Title	
Author	Date
CRFs	Natural Language
Parent links	Child links
Relation links	

Fig. 1: Page structure

3.2 Role of links

Relation links in CCR are all uni-directional links. A common way of organizing knowledge is to create a hierachy in which there is a pyramid structure which provides increasing detailed knowledge as the levels get deeper. This can be achieved using the *relation links*, where the links specified are links to pages with increasing detail. When links are created as such, inference based on these links will draw on knowledge of increasing detail. However, it is also possible to create a network that is more web-like where each page is linked to many peers. A network of pages linked as such would be less structured, but it would still enable the inference engine to gather knowledge from pages that are related to it.

Through the inference described above, it is possible to derive knowledge that has not been explicitly expressed. When the knowledge is derived based on knowledge from two or more pages, the page that forms the basis is considered to be the 'parent' pages and the resultant knowleldge is stored in a new 'child' page. These pages are related to each other by the *parent links* and *child links*.

3.3 Inference through multiple pages

In a system where the knowledge is segregated, it is crucial that when inference is performed, the knowledge used can be expanded to include more knowledge as neccesary. When performing a query against a page, it is possible to expand the amount of knowledge used in the query through the relation links. In figure 2, the example of drawing medical knowledge from another page to provide more details about "appendicitis" is shown. From the additional knowledge that was included in the knowledge base when requested, more information about a certain object can be obtained.

Although this example only shows the grouping of two pages of knowledge, it is possible to perform a multi-level expansion to give rise to inferences that expand the knowledge used as required. All the knowledge in the original page will first be used for the query, then folllowing that all the knowledge from the pages listed in *relation links*. Beyond that all *relation links* of the pages already used will be drawn upon to expand the amount of knowledge. Thus a multi-level expansion of the knowledge can achieved.

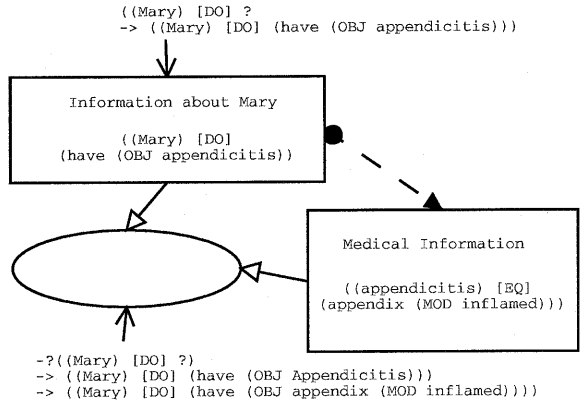

Fig. 2: Multi-page query

4 Conclusion

We have achieved a simple implementation of this meta-structure using HTML-CGI as our user interface. The following figure 3 is a snapshot of the implementation.

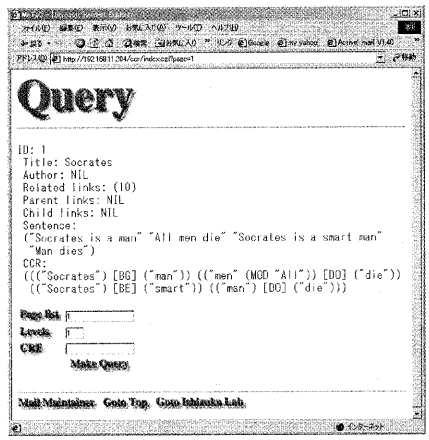

Fig. 3: Query page Snapshot

The Meta-structure for CCR is key component to providing CCR with the means to handle large amounts of data without confusion. It has also been possible to study the interaction between groups of knowledge. In the multi-page query process, we have been able to see how the expansion of knowledge can improve the quality of a query that cannot be satified within a single group of knowledge.

References

[1] 友部　博教, 石塚　満, 木本　協司: "知識表現CCRを用いた知識統合", 人工知能学会知識ベースシステム研究会 (第 50 回)SIG-KBS-A003 (2000).

[2] 友部　博教, 木本　協司, 伊庭　斉志, 石塚　満: "自然言語に近い表現を可能にする概念化学プログラミングの実現", 第 58 回情報処理学会全国大会 (1999).