

## ◆知識情報インフラ

**Acquiring Causal Knowledge from Text Using Connective Markers**

(接続標識に基づく文書集合からの因果関係知識獲得)

乾 孝司 (Inui Takashi)

tinui@lr.pl.titech.ac.jp

奈良先端科学技術大学院大学情報科学研究  
科情報処理学専攻

指導教員：松本 裕治, 乾 健太郎

博士 (工学), 2004 年 3 月 取得

**Keywords** causal relation, knowledge acquisition.

**概要：** A major challenge in natural language understanding is to build a comprehensive common-sense knowledge base in the efficient way possible. This thesis discusses automatic knowledge acquisition from text, especially the acquisition of causal relations. We consider four types of causal relations, namely, cause, effect, precondition and means. They are based on agents' volitionality, as proposed in discourse understanding. The idea behind knowledge acquisition is to use resultative connective markers such as "because", "but" and "if" as linguistic cues. However, there is no guarantee that such a connective marker always signals the same type of causal relation. Therefore, we need to create a computational model that is able to classify samples according to the causal relation. In this work, we focus on Japanese complex sentences including the word *ため* (because). The following questions are asked: (1) What kinds and how much causal knowledge is present in the document collection, (2) How accurately can relation instances be identified, and (3) How can acquired causal knowledge be made available to applications.

First, we investigated the distribution of causal relation instances in Japanese newspaper articles. The main part of this investigation was conducted based on human judgments using linguistic tests. We confirmed that it is possible to acquire causal relation instances from approximately 90% of samples. Second, we assessed how accurately we can automatically acquire causal relation instances by experiments. Using a machine learning technique, we achieved 80% recall with over 95% precision for the cause, precondition and means relations, and 30% recall with 90% precision for the effect relation. Furthermore, the classification results suggest that one can expect to acquire over 27,000 instances of causal relations from one year of Japanese newspaper articles. Third, we applied the acquired causal knowledge to annotate words with its desirability. From this investigation, it became clear that causal relation instances, at least instances of cause relations and means relations, are useful for assigning desirability of words.

**公表論文：** 乾 孝司, 乾健太郎, 松本裕治. 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得, 情報処理学会論文誌, Vol. 45, No. 3, pp 919-933 (2004)

**現職：** 21 世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」PD 研究員 (東京工業大学)

**論文入手先：** <http://cl.naist.jp/>

**抱負：** テキストからの知識獲得など, 自然言語処理の研究を通して, 情報化社会の発展に貢献していきたい。

## ◆知識情報インフラ

**会話エージェントに媒介されたコンテンツ  
マネジメントシステム**

久保田 秀和

kubota@ii.ist.i.kyoto-u.ac.jp

東京大学大学院工学系研究科電子情報工学  
専攻

指導教員：西田 豊明

博士 (工学), 2004 年 3 月 取得

**Keywords:** 会話エージェント, コンテンツマネジメントシステム, テキスト表現・会話的表現変換

**概要：** 本論文の目的は, 個人によるコンテンツ創造を支援するためのコンテンツマネジメントシステム (CMS) の構築である. 従来, 雑多な内容を含む個人コンテンツの発展プロセスを, 計算機を用いて活性化することは困難であった. 本論文ではコンテンツ発展を媒介する存在として会話エージェントを用いることによりこの問題を解決する.

はじめに, 本人の会話を代行する分身エージェントを提案し, EgoChat II システムとして実装する. 分身エージェントは電子メールから抜粋された短文をコンテンツとする会話エージェントであり, 評価実験の結果, その会話が人にとって背景を補うことにより理解可能であったことを示す. 次に, より大きなコンテンツの構築手法として, 1 段落程度のテキストを 2 体のエージェントによる会話的表現へ変換する手法を提案し, 心理学的実験の結果, 本手法が聞き手のテキスト理解を向上させたことを示す. また, 本手法を POCTV システムとして実装し, 1 年間の実証実験を行った結果, 大規模コンテンツを構築可能であったことを示す. 最後に, 個人 CMS を統括する知識チャンネルモデルを提案し, コンテンツ流通の方針を記述可能とするチャンネルポリシー, コンテンツに時系列的な表現を与える会話エージェント, 空間的な表現を与える番組表の三つを設計する. また, 本モデルに基づく CMS を EgoChat III システムおよび番組表生成システムとして実装し, 実コミュニティにおいて運用する. 本論文は以上の実装と実験によって, 個人によるコンテンツ創造を支援する CMS が実現したことを示す.

**公表論文：** 久保田秀和, 黒橋禎夫, 西田豊明: 知識カードを用いた分身エージェント, 電子情報通信学会論文誌「ソフトウェアエージェントとその応用論文特集」, Vol. J86-D-I, No. 8, pp. 600-607 (2003)

**現職：** 京都大学大学院情報科学研究科学術研究支援員

**論文入手先：** <http://www.ii.ist.i.kyoto-u.ac.jp/~kubota/index.html>

**抱負：** コンテンツとは人を表現し, また人と人をつなぐ存在であると考え. 今後, 持続的に発展可能なコンテンツ制作, 利用, 流通技術を確認することによって, 大量のコンテンツに寄り添って生きる現代の個人や家庭, コミュニティに対して貢献したい。