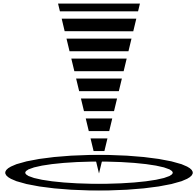


特集



テキスト自動要約

—知的活動支援の基本技術として—

Example 3-1: This problem is not so simple to solve.

Example 3-2: This is a difficult problem.

Example 3-A: This is not so simple to solve.

Example 3-B: This is not an easy problem.

Example 3-C: This is a difficult problem.

Example 3-D: This is a difficult problem.

Example 3-E: This is a difficult problem.

Example 3-F: This is a difficult problem.

Example 3-G: This is a difficult problem.

Example 3-H: This is a difficult problem.

Example 3-I: This is a difficult problem.

Example 3-J: This is a difficult problem.

Example 3-K: This is a difficult problem.

Example 3-L: This is a difficult problem.

Example 3-M: This is a difficult problem.

Example 3-N: This is a difficult problem.

Example 3-O: This is a difficult problem.

Example 3-P: This is a difficult problem.

Example 3-Q: This is a difficult problem.

Example 3-R: This is a difficult problem.

Example 3-S: This is a difficult problem.

Example 3-T: This is a difficult problem.

Example 3-U: This is a difficult problem.

Example 3-V: This is a difficult problem.

Example 3-W: This is a difficult problem.

Example 3-X: This is a difficult problem.

Example 3-Y: This is a difficult problem.

Example 3-Z: This is a difficult problem.

特集：テキスト自動要約 —知的活動支援の基本技術として— 編集にあたって

奥村 学

東京工業大学精密工学研究所
oku@pi.titech.ac.jp

久光 徹

(株)日立製作所中央研究所
hisamitu@harl.hitachi.co.jp

増山 繁

豊橋技術科学大学知識情報工学系
masuyama@tutkie.tut.ac.jp

現在、インターネット等を通じて大量の情報が溢れており、目まぐるしく変化する状況に対応して素早い意思決定を行うためには、大量の情報の中から情報を的確に取捨選択することが必須である。したがって、大量の情報から重要な情報のみを選択して提供し、要点の迅速な把握を支援するテキスト自動要約は、人間の知的活動を支援する基礎技術としてますます重要性を増している。

それに伴い、近年内外で活発な研究がなされてきており、1997年以来、要約をテーマとするワークショップがCOLING, ACL等の自然言語処理に関する国際会議との併設でほぼ毎年開催され、COLING2002では、質問応答システムと要約とをテーマとする併設ワークショップが開催された。また、2003年7月に札幌で開催されるACL2003でも、おなじく、質問応答システムと要約とをテーマとする併設ワークショップが計画されている。また、第2編の記事で紹介するように、米国防総省支援によるTipsterプロジェクトのSUMMACや、同じく、TIDESプロジェクトのDUC (NIST主催)、国内では、国立情報学研究所主催のNTCIRのTSC等の評価型ワークショップが開催されてきている。また、第1編の記事で述べるように、すでにいくつかのワープロソフトやWeb上の検索エンジンにも組み込まれており、身近なものとなりつつある。さらに、たとえば、セマンティックWeb (本誌2002年7月号特集参照)において、必要なときに必要なだけの情報を的確かつ簡潔に提供してくれるような「気の利いた」知的エージェントを実現するための基礎技術として必須であることも見逃せない。

そのような背景の下、本特集では、この分野への関心を高め、研究のより一層の活性化を図るべく、非専門分野の方にもテキスト自動要約技術への興味を持っていただくよう解説する。

第1編では、読者にテキスト自動要約の現状を知っていただくために、複数文書の要約等、重要かつ現在活発

に研究が行われている技術、注目される話題を中心に概説を行った。また、現在どこまで可能になっているかを示すため、Web上で現在公開されている2つのツールを中心にどのように実用化されているかを解説する。

第2編では、自動要約システムの技術を発展させるために必要で、最近の自動要約研究の重要な課題の1つである評価に関して、NTCIRのTSCを中心に評価型ワークショップの目的と意義、実施内容、評価方法、得られた知見、今後の展望等について、タスクオーガナイザとしての経験を踏まえて解説する。

第3編では、まず、携帯電話、PDA等の携帯端末へのテキスト配信に関する問題をハードと通信の環境、利用局面、目的、応用の4つの局面から論じる。次に、インタラクティブなブラウザについて紹介する。また、携帯端末へのテキスト配信のための要約技術として、生成型の統計的要約手法、抽出、編集型の要約手法を紹介する。さらに、縮約(言い換え)による方法を例文を用いて分かりやすく紹介する。

第4編では、通信・放送機構での研究を中心に、聴覚障害者のための字幕放送のための要約について紹介する。まず、現在の字幕についての分析を示し、次に、手法として、書き換えによる方法、文節圧縮法について紹介し、さらに、ニュース、教育・ドキュメンタリー、ドラマ・バラエティ等の分野ごとの可能性について論じる。

第5編では、将来の社会における知的活動支援のための基本的要素技術としての観点から、テキスト自動要約の要素技術として重要な統計的方法と換言処理について紹介し、さらに、これからテキスト自動要約が必要とされる応用分野、および、研究を進める必要性が大きい基礎技術について展望する。

末筆ではあるが、ご多用にもかかわらず、快く記事を寄せていただいた著者の方々、適切なコメントをいただいた読者の方々に厚くお礼申し上げます。

(平成14年11月5日)