

異種知識の協調利用に基づく情報検索システム

3F-4

羽生田博美(haniuda@rwcp.or.jp) 木下哲男

新情報処理開発機構 新機能沖研究室

1. はじめに

今日、電子郵便・WWW・ネットワークニュース等に見られるように、膨大な量の情報が世界的規模で日々発生し、その規模の拡大・種類の多様化はますます進んでいる。このような状況下では、情報の利用者は専門家から非専門家へと移行せざるをえず、利用者がしばしば感じるのは、自分が望む情報を既存のシステムでは思うように得られないというもどかしさである。

筆者らはこれを実世界における問題解決システムとして解決すべき問題と考え、それに必要なやわらかな情報処理機能の検討を進めている。本稿では、こうした情報検索システムの基本的枠組を提案する。

2. 情報検索に必要とされる機能

冒頭で述べたような状況下において求める情報を思うように得ることが出来ない理由としては主に2つが考えられる。まず、ネットワークニュースが端的な例であるが、情報の構造化がなされていない。これがハイパーテキスト構造などを導入しても解決されないことは、WWWなどを利用してみれば明らかであろう。迷子の問題と同時に、利用者が必要とする情報を的確に表現することが一層難しくなってくるからで、これが第2の理由である。検索対象が膨大・多様になると同時に利用者が非専門家化するために、情報フィルタリングと同様の問題[森田93]が生じるのである。

これらの問題に対する様々なアプローチがある。例えば、単語の共起関係に基づいて情報を分類して構造化する[有田93]、単語の出現頻度などをもとに作成したベクトルによって情報を表現するベクトル空間モデルやその拡張（[Foltz92]のLatent Semantic Indexingなど）、個々のキーワードに競合エージェントを割り当てて利用者の評価からその興味関心の

記述を獲得する[Bac92]、シソーラスにより質問を拡張することで質問をよりの確なものにしようというアプローチ等々である。

しかし、どのアプローチにおいても一定の構造を利用者に強制することになり、冒頭述べたような状況では適切な構造が利用者の要求によって変化しうするため、十分に対応することは難しい。例えば、シソーラスはある一定の上位・下位・関連語関係などに構造化されているため、検索対象の持つ意味内容のすべてを反映することはできないからである（例えば[菅野94]では、シソーラス検索とフリーキーワード検索が相補的であることが指摘されている）。

これらの問題に対処するためには、利用者の質問に対してその意図するところを正確かつ十分につかむと同時に、あらかじめ組織化された構造に適しない質問や情報構造の変化に耐えて適切な回答を返す機能が必要とされる。すなわち、知性・恒常性・発展性といった特性＝やわらかな情報処理能力[菅原94]が必要とされる。

3. 基本モデル

前節で例示した様々なアプローチは、それぞれが個別の知識（例えば文書間での単語の共起関係、キーワードの上位・下位関係等々）に基づいて特殊化されているとも言える。したがって、やわらかなシステムとなるためには、まず個々の異種知識を十分に利用するためにそれぞれの知識に特殊化された処理機構が必要とされる。これによって、質問や情報構造の変化の影響を局所化することも可能になり、ある程度の知性や恒常性をシステムに与えることができる。また、各処理機構は、学習や自己組織化によって環境の変化への適応能力を持つ必要がある。そして、これらの処理機構は、個々の知識に関して特殊化されている、すなわち相補的であるため、各処理機構間での異種知識の協調的利用が必要不可欠な要素となる。図1に、本稿で提案する方式に基づく、ネットワークニュース記事を対象とした情報検索システムの基本モデルを示す。

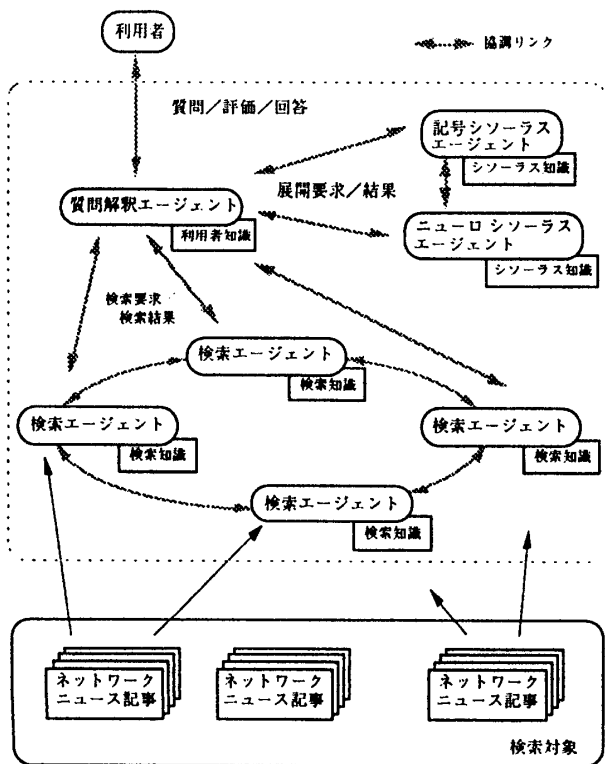


図1 協調型情報検索システム

システム全体は、シソーラス・検索・質問解釈等のエージェント群（およびそれぞれのバリエーション）から構成されるマルチエージェントシステムである。各エージェントは、例えば、シソーラスエージェントは辞書から抽出した単語間の意味関係や検索対象中の単語の共起関係を学習により獲得したシソーラスを利用した検索語の展開、各検索エージェントは異なる戦略に基づく検索処理、質問解釈エージェントは利用者の検索履歴や質問結果に対する評価などに基づく回答の生成などを協調的に行うといったように特殊化されている。また同一種類の処理を行うエージェントも、例えば記号シソーラスは単語間の関係を個々別々に扱うことを得意とし、ニューロシソーラスはそれらを全体的かつ連続的に扱うことを得意とするといったように特殊化され、協調的なエージェントとして利用される[奥村94]。

一方、関連フィードバックを用いて利用者の興味関心に関する記述を得る方式は、[Foltz92, Bac92]のように文書間を通じての語の使用パターンにおける潜在的な構造を推定して利用するものであった。我々は、潜在的情報を利用するこうしたアプローチを、異種知識の協調利用という枠組で拡張する。すなわ

ち、シソーラス展開処理は、利用者の質問履歴および回答に対する評価により制御する。また、検索エージェントの検索結果から抽出した単語によりシソーラス展開を制御可能とするが、その場合にも何らかの分類（例えば[佐々木95]）を行いその結果を利用することも考慮する。すなわち、本稿で提案する方式では、それぞれのエージェントが顕在的・潜在的情報を利用するとともに、それぞれのエージェントだけでは捉え切れない情報を他のエージェントとの協調を通じて利用するというモデルとなっている。

4. おわりに

異種知識を協調的に利用することによりやわらかな情報検索システムを実現する枠組について述べた。現在、各種エージェントを試作すると共にそれら間における協調で重要な役割をはたす特徴量を検討するため実験を行っている[奥村95, 佐々木95, 城風95]。今後は、情報検索システムをテストベッドとして異種知識の協調利用の枠組を確立することを通じて、実世界の問題解決に必要な異種知識利用の基本的枠組を探っていく予定である。

参考文献

- [Bac92] P. E. Baclace, "Competitive Agents for Information Filtering," CACM 35-12, 1992.
- [Foltz92] P. W. Foltz, S. T. Dumais, "Personalized Information Delivery: An Analysis of Information-Filtering Methods," CACM 35-12, 1992.
- [有田93] 有田, 「AIにおける問題解決技術の新展開—自己組織化情報ベース—」, 電気関係学会東海支部連合大会, 1993.
- [奥村94] 奥村, 木下, 「情報検索用シソーラスの自己組織化について」, 信学技報AI94-47, 1994.
- [奥村95] 奥村他, 「情報検索用自己組織型シソーラスの評価」, 第50回情処全国大会3F-1, 1995.
- [佐々木95] 佐々木他, 「SCONNによる文書の自動分類」, 第50回情処全国大会3F-3, 1995.
- [城風95] 城風他, 「種々のキーワード間類似度を用いたネットワーク記事検索」, 第50回情処全国大会3F-2, 1995.
- [菅原94] 菅原, 木下, 白鳥, 「やわらかいシステムの形式化に関する一考察」, 信学技報AI93-84, 1994.
- [菅野94] 菅野, 草田, 「MEDLINEのCD-ROM版におけるフリーキーワード検索の検討」, 情処研報94-FI-35-2, 1994.
- [森田93] 森田昌宏, 「情報フィルタリング技術の現状と展望」, 信学技報AI93-24, 1993.