

情報検索エージェントを用いた グループ発想支援システムの試作について

水口卓也 伊藤孝行 新谷虎松

名古屋工業大学

E-mail: mizuguti@ics.nitech.ac.jp

1. はじめに

会議の負担を軽減するためにグループウェアのニーズが高まっている。会議は、伝達会議、決定会議、調整会議、創造会議に種類分けされ[1]、創造会議が最も重要な会議である。そのような背景から特に創造会議を支援するグループウェアが求められている。創造会議は、発散段階と収束段階に分けられる。発散段階は、議題の解決のためのアイデアを出す段階である。会議の参加者は、良し悪しにかかわらず出来るだけ多くのアイデアを出し合う。文献検索などをはじめとする情報収集手段を用いてデータ収集を行い、アイデアを出すことが重要である[2]。収束段階は、出されたアイデアを整理して評価して具体策とする段階である。

WWWが普及したことで多くの情報が簡単に手にはいるようになった。これらの情報を効率よく効果的に利用するための研究も盛んに行われている。

本研究では、WWWで得られるデータを創造会議の発散課程に利用して、会議の参加者の発想を促すことを考える。本研究で試作したシステムの特徴は、会議を監視するエージェントがアイデアがある程度出尽くしたことを判断すると、自動的に議題に関連があるであろうキーワードをWWWから探してきて会議の参加者に提示する点である。本稿では、試作したシステムの概要と、システムの構築に用いたエージェントの動作について述べる。

2. WWWに基づく発想支援システム

本システムは、遠隔地・同期型のグループウェアである。会議の参加者は、インターネットに接続された端末を用いて、遠隔地からサーバに接続することで会議に参加する。会議の概念図を図1に示す。各端末ではコミュニケーションツール(チャットウィンドウなど)、エージェントツール、WWWブラウザ

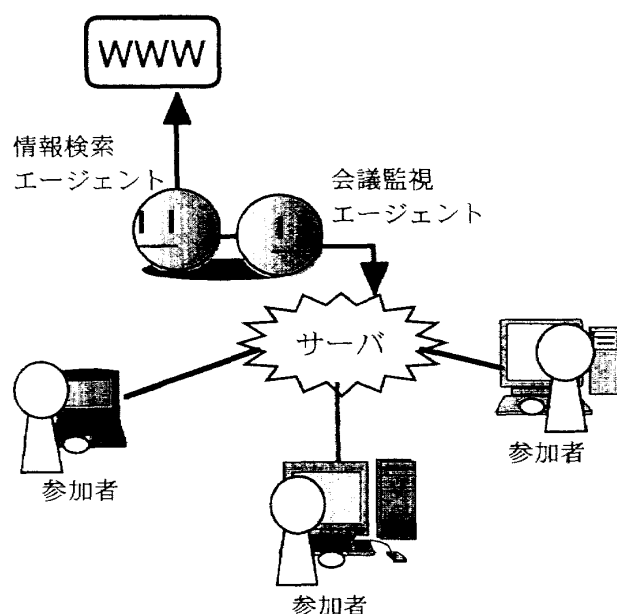


図1 会議の概念図

などが実行されている。コミュニケーションツールは文字などで会議の参加者同士が通信を行う。エージェントツールはサーバ上のエージェントと通信を行う。WWWブラウザは情報の閲覧に用いる。サーバ上には会議監視エージェントと情報検索エージェントが存在する。会議監視エージェントは、参加者のコミュニケーションツールの内容を監視し、最近の会話の中からキーワードを抽出して出現頻度を計算する。コミュニケーションツールで交される会話が少なくなってきたら、出現頻度の高いキーワードを情報収集エージェントに渡す。情報検索エージェントはキーワードが渡されると、そのキーワードを検索エンジンで検索する。得られたページの集合から適当なものを選び、キーワードを抽出する。ページから抽出したキーワードのうち、会議監視エージェントが抽出したキーワードではないものがエージェントツールを通じて各参加者に提示される。会議の参加者は情報検索エージェントによって提示されたキーワードを元に再びアイデアを出す。

3. 会議監視エージェント

コミュニケーションツールはWYSIWIS(What You See is What I See.)が守られており、各参加者間で表示が異なることはない。会議監視エージェントは、コミュニケーションツールの内容を監視する。監視する対象は、チャットウィンドウ(文字による会話)だけとする。会議監視エージェントは、最近の発言M個の中からキーワードを抽出し、その出現頻度を計算する。抽出されたキーワードの集合 $K(k_1, k_2, \dots, k_p, \dots, k_n)$ の各要素の出現回数を $N(k_i)$ で表わすと、出現頻度 p_i は、

$$p_i = \frac{N(k_i)}{\sum_{j=0}^n N(k_j)}$$

で表わされる。出現頻度が高いキーワードが現在の議論の内容に関係あるキーワードであると判断する。

発言間隔が長くなると、議論が一段落してアイデアが出にくくなったと判断し、会議監視エージェントは出現頻度の高いキーワードを情報検索エージェントに伝える。情報検索エージェントは、WWWから得た新しいキーワードを参加者に提示する。新しいキーワードを参加者に提示することで発想を促し、より多くのアイデアを集めるようにする。

4. 情報検索エージェント

情報収集エージェントは、渡された検索キーワードを元にinfoseek(<http://www.infoseek.co.jp/>)やgoo(<http://www.goo.ne.jp/>)などの検索エンジンを用いて、検索キーワードに関連のあるページを検索する。検索して得られたページの集合から、新しいキーワードを会議の参加者に提示するまでの流れを図2に示す。まず最初に、得られたページの集合から無駄なものを取り除く。リンクが無効になっているもの、複数個存在するもの、ページのプレビュー(検索エンジンで検索すると得られる)に検索キーワードが存在しないものを取り除く。次に、それぞれのページのキーワードを抽出し、tf・idf[3]を用いて重みづけを行う。この重みづけは他のシステムでもよく用いられる。キーワード t_i のホームページ d_j に対する重みTFIDF(t_i, d_j)は、

$$TFIDF(t_i, d_j) = FREQ_{t_i, d_j} \cdot \{1 + \log_2(N / DFREQ_{t_i})\}$$

で表わされる。 $FREQ_{t_i, d_j}$ はキーワード t_i のページ d_j における出現頻度、 N は全ページ数、 $DFREQ_{t_i}$ は

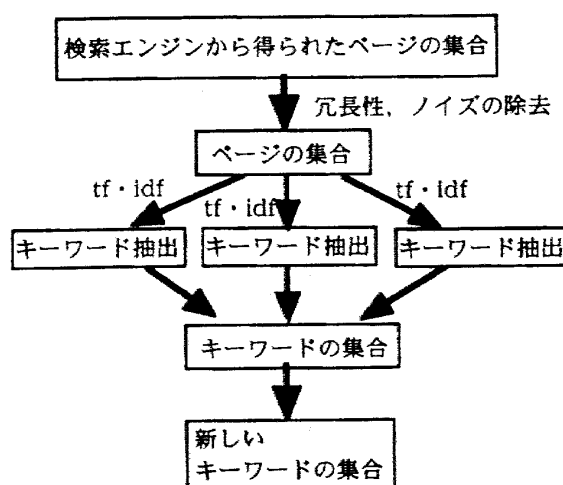


図2 情報検索エージェントの作業の流れ

t_i が出現するページ数である。

tf・idfを用いれば、そのページの特徴となるキーワードが得られる。ページごとに各キーワードの重みを計算し、重みの大きいものから適当な数のキーワードを取り、ページから得られたキーワードとする。ページの集合の要素全てから上記の方法でキーワードを抽出し、キーワードの集合を作る。キーワードの集合から会議監視エージェントが発言から抽出したキーワードを削除し、新しいキーワードの集合とする。

5. おわりに

会議を監視するエージェントが、自動的に議題に関連があるであろうキーワードをWWWから探してきて会議の参加者に提示することで創造会議の発散課程を支援するシステムの構築について述べた。

今後の課題としては、発言に現われるキーワードをそのまま検索に用いるのではなく、概念辞書などを用いて類似語を判断し、発言の流れをより正確にとらえて検索を行うシステムなどが考えられる。

参考文献

- [1] 高橋 誠, "会議の進め方", 日本経済新聞社, 1987.
- [2] 国藤 進, "グループで発想する", 知的触発に向かう情報社会, 共立出版, pp.30-41, 1995.
- [3] G. Salton, "Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer", Addison Wesley, 1989.