

論文の内容の要旨
感性・概念による情報検索技術に関する研究
氏名：川前徳章

本論文は、「感性・概念による情報検索技術に関する研究」と題し、七章より構成されている。近年、インターネットやPCの普及に伴って、我々は多くの電子化された情報や資料をオンラインで入手することが可能となった。その代表的なものとしてwww(world wide web)からの情報の入手があり、必要な情報の入手を支援するサービスに検索システムがある。今後も効率的に必要な情報を入手するために、検索システムの重要性はますます高まっていくと考えられるが、現在の検索システムは必ずしも情報の入手が効率的ではない。

その原因は、ユーザが情報の検索を目的としているのに対し、現在の検索システムがデータベース検索技術の転用に過ぎず、情報でなくデータの検索になっていることにある。本論文は、ユーザにとって情報検索が効率的に行えることを目的として、この検索システムに“人のコミュニケーション”を導入することを提案する。“人のコミュニケーション”を導入することで、検索システムは従来の単なるデータベース技術の応用から、感性と概念による情報検索が実現できるようになる。この結果、本論文で提案する検索システムは、ユーザのキーワード入力部分においては適切なキーワードを提示すること、検索結果においては内容の類似性によって検索結果を分類して表示することを行っている。以下、各章の構成を紹介する。

第一章は「序論」であり、従来の情報検索の概要とその問題点を自然言語の利用と検索方式の立場から述べている。次にこれらの問題点を踏まえて、情報検索がユーザにとって効率的になるためにユーザ側の支援と検索結果の改善が必要なことを明らかにし、“人のコミュニケーション”に着目しこれを検索システムに導入することを提案している。更に本論文で提案した検索システムの成果をまとめ、最後に本論文の構成を示している。

第二章は、「既存研究」と題し、これまでに提案されてきた情報検索技術をユーザ側と検索結果に類別し、それぞれにおいてその非効率性を示している。その原因が、従来技術には“人のコミュニケーション”の観点が入力されておらず、情報の検索でなくデータの検索に留まっている事を結論付けている。

第三章は、「ユーザの検索行動に基づいたキーワード提示」と題し、“人のコミュニケーション”を導入したキーワード提示手法を紹介している。提案手法は、ユーザがキーワードを検索要求として入力すると、そのキーワードの検索結果の表示に加え、入力されたキーワードと関連性の高い順に別なキーワードの提示を行う。関連性はユーザの再検

索でキーワード変更するときの検索行動を、“別表現・具体化・抽象化・同概念層・ニーズ変化”と五種類に定義し、その行動別におけるキーワードの関係のことである。検索システムは、関連性が高いキーワードほど高い確率でユーザがその検索行動で次の検索要求として利用すると予測して提示を行う。例えば、ユーザが検索要求として「りんご」を入力した場合、別表現ならば「林檎」、具体的ならば「ゴールド」などというキーワードの提示を行う。提示されるキーワードの数は、ユーザが任意に決定することができる。従来はユーザがこれらのキーワードを自力で発見することが要求されていたが、検索システムがキーワードを提示することで、ユーザは自力で探す負担が軽減され、検索行動の効率化が実現される。手法の実現のために本論文はユーザの検索履歴を利用し、ユーザの検索行動のモデル化を行う。このモデルによって、検索履歴からキーワード間の関連性を求めることができ、検索行動別にユーザに対して適切なキーワード提示が実現できるようになる。

第四章は、「ユーザの検索行動別のキーワード提示」と題し、第三章で提案したキーワード提示手法の適用について論じている。ここでは、実際の検索システムの検索ログに提案したキーワード提示手法を適用した結果、既存研究では支援できなかったユーザや検索行動までが支援できることを示し、“人のコミュニケーション”における情報要求の具体化を実現していることを確認した。

第五章は「単語からノイズを除去した文書分類」と題し、人のコミュニケーションを導入した文書分類手法を紹介している。提案手法は、従来はユーザが入力したキーワードを含む文書のインデックスのリストに過ぎなかった検索結果を、内容の類似性によって構造化して表示する。この検索結果にはユーザが入力したキーワードは含まれなくても、内容が類似した文書のインデックスが含まれる。例えば、ユーザが検索要求としてキーワードを「統計」と入力した場合、検索結果は「統計」に関する本、アプリケーション、シラバス、解説文書などに自動的に分類され、解説文書の分類には「統計」というキーワードが含まれなくても「統計」に関わりの深い文書までが含まれる。従来はキーワードを含む文書を検索するキーワードマッチングであったが、提案する手法によってキーワードを含まなくても内容が類似した文書の検索が実現できる。手法の実現のために本論文は照応解析の導入、分類する文書内の単語の重みのつけ方、単語の選択方法、因子分析の導入、要約の作成の提案を行った。更に因子分析においては、新しい因子得点の計算法と次元数を決定する評価式の導出を行っている。これらの提案の成果は、従来よりも少数の単語から計算された文書の類似度が我々の判断する内容の類似度に近くなることが実現できる。

第六章は「情報論的アプローチによる文書の分類」と題し、第五章で提案した文書分類手法の適用について論じている。ここでは実際のニュース記事に提案手法を適用した結果、既存研究よりも文書の分類結果が我々の判断する内容の類似性に近くなったことを示し、

人とのコミュニケーションにおける内容の判断を実現していることを確認した。

第七章は「まとめ」であり、本論文の研究成果をまとめ、残された課題や今後の研究の方向性について整理している。

以上、本論文では、情報検索がユーザにとって効率的になるために、人のコミュニケーションを導入した検索システムの提案を行った。情報検索が効率的になるためにはユーザ側の支援と検索結果の改善が必要になる。提案手法は、ユーザ側の支援として、ユーザがキーワードを検索要求として入力すると、そのキーワードの検索結果の表示に加え、入力されたキーワードと関連性の高い順にキーワードの提示を行う。また検索結果においては内容の類似性によって構造化して表示する。この検索結果にはユーザが入力したキーワードは含まれなくても、内容が類似した文書のインデックスが含まれる。その結果、これらの提案手法が人とのコミュニケーションにおける「情報要求の具体化」と「内容の判断」を実現することにより、感性と概念による情報検索が実現できて、ユーザにとって効率的になったことが明らかになった。