

# ナレッジマネジメントに関する研究

電子メール・電子会議室からのナレッジ抽出機能の開発

## Study on Knowledge Management

Development of a Function to Extract Knowledge from E-mails and Electronic Forums

(電力技術研究所 お客さまネットワークG 情報通信T)

現在、日常業務において電子メール・電子会議室等のシステムが頻繁に利用されている。これらのシステムには、日々の業務上の知識が蓄積されていると考えられる。今回、電子メールと電子会議室のデータから、Q&A抽出・質問者に関する属性情報の統計表示・質問カテゴリに関する属性情報の統計表示という形で知識抽出を行う機能を開発したので報告する。

### 1 研究の背景と目的

ITの普及により、企業において電子メール・電子会議室等のシステムが日常的に業務利用されている。これらのシステムには多くの情報が蓄積されるが、最新の情報にしか目を通されない傾向にある。しかし、システム内でやり取りされた電子メール等の情報には、共有すべき知識が蓄積されていると考えられる。例えば、業務内容で頻繁に直面する問題の解決方法や、過去やり取りしたQ&A等があり、これらの埋もれた知識を共有できれば、業務効率の向上につながると考えられる。ただし、埋もれた知識はそのままでは利用することができないため、必要な時に必要な情報を分かりやすい形で表示する必要がある。

そこで、本研究ではシステムに蓄積されている情報から知識を抽出し、共有する手法を開発する。具体的には、電子メール・電子会議室のデータから、

- Q&A集の抽出(電子メール・電子会議室)
- 質問者に関する属性情報の統計表示(電子メール)
- 質問カテゴリに関する属性情報の統計表示(電子会議室)

という形で知識抽出を行う試作システム(以下、「本システム」という。)を開発する。これにより、必要な情報の抽出・表示を行うことができる。

### 2 システムの概要

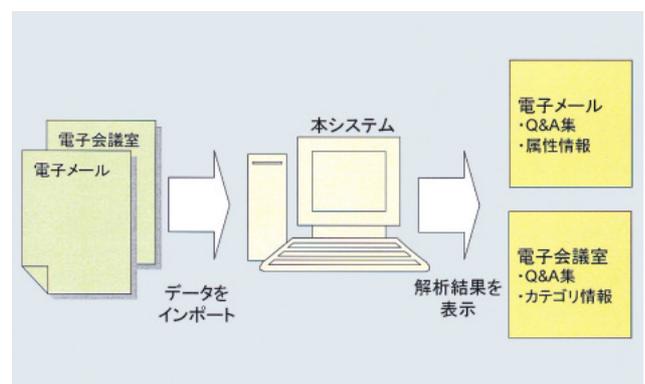
本システムの概要を示す(第1図)。

電子メールと電子会議室のデータを本システムにインポートしデータの解析を行う。電子メールのデータからは、Q&A集の抽出と、質問者の入力した属性情報の集計を行う。また、電子会議室のデータからは、Q&A集の抽出と、投稿カテゴリ集計を行う。

なお、これらの解析結果を共有するために、本システムはWebアプリケーション形式で構築する。

(Information and Communication Team, Customer Supply Network Group, Electric Power Research and Development Center)

E-mail and electronic forums are heavily used in modern day-to-day business. It is conceivable that knowledge needed for daily business has accumulated in these systems. A system has been developed to perform knowledge extraction from data stored in e-mails and electronic forums. The system extracts knowledge in the form of Q&A extraction, statistical display of attribute information concerning the questioners, and statistical display of attribute information concerning the question categories. The system will be introduced in the following.



第1図 システム概要

### 3 Q&A抽出

本システムでは、以下2種類のデータを想定している。

電子メールデータ：質問者は問い合わせフォームに質問事項、メールアドレス等を入力し送信する。質問は電子メール形式で送信される。回答者は質問に対する回答を電子メールで返信する。

電子会議室データ：質問者は質問のカテゴリを選択し、質問事項を入力し投稿する。回答者は質問投稿に対する返信として、回答を入力し投稿する。

電子メールデータでは、問い合わせフォームの指定の入力欄に質問が入力されているため、そこに入力された文面を抽出し、挨拶文・締め括り文に相当するものがあれば除外して質問文とする。また、回答については、返信メールの文面を抽出し、挨拶文・締め括り文・署名等の、回答以外の部分を除外して回答文とする(第2図)。なお、挨拶文・締め括り文の判定は、それらの文に含まれる文言をあらかじめ設定しておき、メール文面と照合することで行う。

そして、質問メール・回答メール双方からメールアドレスと送信時間を抽出し、質問文と回答文の紐付けを行う。

質問 花子様 この度は、弊社ホームページをご利用いただきありがとうございます。質問者様からのお問い合わせにつきまして、次のとおりにお返事させていただきます。	挨拶文 (除外)
弊社の会社情報につきましては、中部電力ホームページの「企業情報」に掲載しております。 また、資料につきましては「資料請求フォーム」をご用意しております。	回答文 回答文
今後とも弊社事業にご理解とご支援を賜りますようお願い申し上げます。	締め括り文 (除外)
中部電力株式会社 ○○部 □□G 中電 太郎	署名 (除外)

第2図 回答文の判定

電子会議室データでも電子メールデータと同様に、投稿された文面を抽出し、挨拶文・締め括り文・署名等の、質問・回答以外の部分を除外して質問文・回答文とする。

そして、電子会議室のツリー構造から質問文と回答文の紐付けを行う(第3図)。

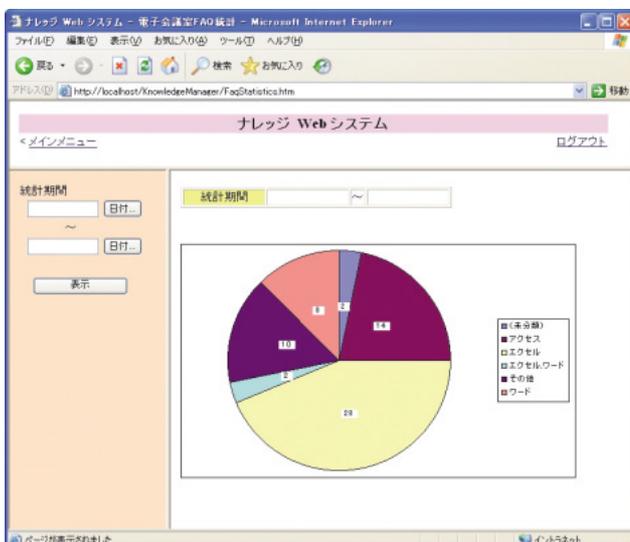
日付	件名	質問	回答	カテゴリ
2008/01/07	Q エクセルのプロパティについて(氏1 名1)			
2008/01/08	T ----- A Re: エクセルのプロパティについて(S_HD マスター)			
2008/01/04	Q エクセル 条件付き書式の格納について(氏2 名2)			エクセル
2008/01/08	T ----- A Re: エクセル 条件付き書式の格納について(S_HD マスター)			
2007/12/20	Q 社算式内の街区物についてほか(氏3 名3)			エクセル
2007/12/25	T ----- A Re: 社算式内の街区物についてほか(S_HD マスター)			
2007/12/14	Q アクセスからエクセルヘッダーを参照する。(氏4 名4)			アクセス
2007/12/14	T ----- A Re: アクセスからエクセルヘッダーを参照する。(S_HD マスター)			
2007/12/13	Q 一部のグラフがカラー印刷できない(氏5 名5)			エクセル、ワード
2007/12/14	T ----- A Re: 一部のグラフがカラー印刷できない(S_HD マスター)			
2007/12/13	Q フォントエディター(氏6 名6)			その他
2007/12/14	T ----- A Re: フォントエディター(氏6 名6)			
2007/12/12	Q フォントエディター(氏6 名6)			その他
2007/12/13	T ----- A Re: フォントエディター(S_HD マスター)			
2007/12/11	Q プロパティの格納について(氏7 名7)			エクセル
2007/12/12	T ----- A Re: プロパティの格納について(S_HD マスター)			
2007/12/10	Q ワードファイルが削除出来ません。(氏9 名9)			ワード
2007/12/11	T ----- A Re: ワードファイルが削除出来ません。(S_HD マスター)			

第3図 電子会議室のツリー構造

## 4 属性情報の表示

問い合わせフォームには、年齢・職業・性別等の属性を入力する項目がある。ここに入力されたデータから、質問者の属性を集計し円グラフで表示する。

電子会議室には、投稿者が投稿内容のカテゴリを指定できる。このデータから、各カテゴリへの投稿数を集計し円グラフで表示する(第4図)。



第4図 属性情報(カテゴリ)の表示

## 5 ユーザインタフェース

ユーザは、電子メールや電子会議室のデータ登録・Q&A一覧の表示・個別のQ&A詳細表示・属性の統計情報表示を、Webブラウザで行う(第5図)。



第5図 ユーザインタフェース

## 6 今後の展開

本研究では、電子メール・電子会議室に蓄積されたQ&A情報、属性情報を抽出し表示する試作システムを開発した。本システムは、業務を通じてシステムに蓄えられたナレッジの整理・共有をするために有効であると考えられる。

今後の課題としては、

質問文・回答文の抽出精度の向上

Q&Aの分類

Q&Aのランク付け

Q&Aの傾向分析

が挙げられる。

については、本研究ではあらかじめ文言を設定しておくことで挨拶文・締め括り文の判定を行っていたが、固定的な表現に依存しない判定方法を確立することが必要であると考えられる。

については、類似のQ&Aはまとめる・利用者によるタグ付け機能を追加する等の改善が必要であると考えられる。

については、利用者による評価機能を追加し、ランクの高いQ&Aは優先的に表示させる等の改善が必要であると考えられる。

については、単語の出現頻度を時間軸で表示する・質問者の属性情報を軸に表示する等の改善が必要であると考えられる。

最後に、本システムの適用シーンとしては、(1)電子メールを用いたお客さま対応におけるナレッジの整理と共有、(2)社内における業務上のQ&Aのやり取りを通じて蓄積されるナレッジの整理と共有、等の利用を想定している。なお、本システムは試行段階であり、今後もナレッジの蓄積・整理・共有に有効なシステムとして各部門に採用を働きかけていく予定である。



執筆者 / 福田 篤