知能と複雑系 113−11
(1998. 7. 24)

インフォーマル情報の自動分類に関する検討

爰川 知宏†, 岩木 英明††, 稲田 善明†, 萱野 忠†

†NTTマルチメディアネットワーク研究所
〒239-0847 神奈川県横須賀市光の丘 1-1-523A

††名古屋大学 工学部
〒464-8603 名古屋市千種区不老町

あらまし 組織活動におけるインフォーマル情報は，業務を円滑化するノウハウや，知的触発の契機となる有用なものも多く含まれている．これらの情報を効率的に蓄積・共有するために，情報組織化手法を検討し，要求条件の整理を行った．情報組織化のためのパラメータを抽出し，そのうち情報カテゴリについて自動分類のためのアルゴリズムの提案と評価を行い，カテゴリ分類上のキーワード抽出の問題を明らかにした．また，これらの複数の情報組織化パラメータを用いた情報組織化システム KnowHow Navigator の提案と試作を行った．

和文キーワード 情報検索，グループウェア，キーワード抽出，情報共有

Organization Technology of Informal Information

Tomohiro KOKOGAWA†, Hideaki IWAKI††, Yoshiaki INADA†, and Tadashi KAYANO†

†NTT Multimedia Networks Laboratories
e-mail: {koko, inada, kayano}@nttmhs.tnl.ntt.co.jp
1-1-523A Hikarinooka, Yokosuka, Kanagawa, 239-0847 JAPAN

††Faculty of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-City, 464-8603 JAPAN
e-mail: hideaki@bioele.nuee.nagoya-u.ac.jp

Abstract Sharing of informal information such as know how and ideas is important to support whole activity in an organization. To support gathering and sharing informal information, we analyzed the fundamental requirements of information organization and classified organization parameters. In these parameters, we focused on an information category and suggested an algorithm of making category lists. By evaluating this, we found some serious problems of picking up suitable keywords from informal information. We also suggested a system about informal information organization system "KnowHow Navigator" which considers the organizational parameter classified above.

Keywords Information Retrieval, Groupware, Keywords Extraction, Information Sharing

1. はじめに

インターネット技術の発展と普及により，オフィスをとりまく情報環境も急速に変化しつつある．電子メール，電子ニュース，WWWなど，インターネットサービスを軸とした組織内情報環境の構築（イントラネット）や，市販グループウェアの導入などにより，単なる情報の電子化にとどまらない，組織内での効率的な情報の蓄積／共有環境の整備は重要な問題と認識されつつある．

グループワークにおいては，会話や過去のノウハウ，さらには作業過程にて発生した新たなプロセスなど，断片的でかつ，暗黙のうちに共有される情報は多い．組織の能力を高め，業務を効率的に進めるために，これらインフォーマル情報をグループ内で効率的に蓄積／共有することが重要である．一方で，これらの情報は重要性が直接的に認識されにくく，情報自体が断片的なため，その情報が活用された場面における暗黙の前提条件を多く含むため，定型化や再利用が困難であるという問題がある．したがって，蓄積／共有の過程で欠落する情報発信者の意図や暗黙の前提を補完するアプローチが必要である．そのために，情報相互の関連づけや，情報分類といった情報組織化のアプローチが必要である．

本研究では，組織におけるインフォーマルな情報に対して，情報共有サービスを介した共有を行う上で問題となる情報組織化の要求条件をまとめる．その中で大きな問題であるカテゴリ分類の実装評価を行い，システム構築による情報組織化へのアプローチを示す．

2. 情報組織化の検討

2.1 インフォーマル情報共有の課題

本論文では，「インフォーマル情報」という用語を，組織活動全般にわたり様々な場面で断片的に発生し，暗黙のうちに共有される情報と位置付ける．会話や過去のノウハウといった，組織活動のさまざまな場面で現れるものに加え，個人に内在したノウハウ，アイデアといった，組織で共有することで業務を円滑に進められるような情報も含めて扱う．また，これらの情報は失われやすいことから[1]，蓄積・共有という観点から，これらの情報を効率的に蓄積し，加工・再利用可能な状態にして扱うための枠組みも検討すべき対象である．

インフォーマル情報は，属人的，断片的，かつ単発的といった特徴のため[2]，

(1) その情報単体では意味が不十分であり，その情報が発せられた前後のコンテキストを考慮する必要がある．

(2) 情報が体系立てられていないため，情報の網羅的な分類が困難である．

といった問題がある．そのため，単なる情報検索サービスの構築では不十分であり，情報の補完アプローチが必要である．

インターネット技術の発展と普及により，世界中から必要な情報を取得するのが容易になった．従来から用いられている電子メールやネットニュースに加え，WWWを用いた情報発信が組織レベル／個人レベル共に非常に盛んに行われている．また，WWW上に構築された膨大な情報から必要なものを容易に見つけだせるようにするために，商用の広域検索エンジンが広く利用されている．

インターネット上で用いられているこれらの情報共有技術は組織情報共有においても適用できるものが多いが，インターネット上の情報と組織内のインフォーマル情報には以下のような差異がある．

(1) 情報発信者・受信者ともに同じ組織内にいることから，面識のない情報発信者による一方的な情報発信でなく，相互のコミュニケーション過程で得られた情報が多く含まれている．

(2) 情報の絶対量はインターネット上に比べてはるかに少ないが，内容的には再利用性が高いものが多い．

(3) 業務ノウハウなど組織固有の情報が多く含まれる．

したがって，インフォーマル情報の共有を支援するためには，

(1) 組織メンバ間のコミュニケーションの促進や，情報発信の負荷低減により，情報発信しやすい環境を作る．

(2) 少ない情報を効率的に再利用できるように情報検索精度を上げる，適切な分類を行う，などを行う．

(3) 組織固有の状況に応じた情報提供（検索ルール，分類方法など）を行う．

などを考慮する必要がある．

2.2 情報組織化のアプローチ

前節で述べたように，組織におけるインフォーマル情報の共有には，情報の再利用性を

高めることと，そのために組織固有の状況に応じた情報提供が必要である．そのために，断片的に生成されたインフォーマル情報を，組織の状況や内容により分類し，情報間を相互に関連づけるアプローチ（情報組織化）が必要である．

情報組織化の課題は以下の２点である．

(1) 情報組織化の基準（何をもって情報間に関連があるといえるか）

(2) 情報組織化の表現方法（情報間の関連をどのように可視化するか）

以下の章ではこの２つの課題に対する要求条件の抽出と支援アプローチについて検討する．

3. 情報組織化方式の検討

3.1 情報組織化のためのパラメータ

インフォーマル情報の組織化は，その情報が内在する意図や前提情報を補完することを目的としている．これらの情報を得るには，その情報が生成された前後のコンテキストや状況に基づく情報分類が必要である．

情報生成時の状況を表すパラメータとしては，以下のものが考えられる．

(1) いつ生成されたか（情報生成時期）

(2) どこで生成されたか（情報生成場所）

(3) 誰が（あるいはどの集団が）生成したか（情報生成者）

(4) どのような内容のものか（情報カテゴリ）

一方で，組織内でこれらの情報が再利用されていく過程で，情報の利用状況が組織内で定着し，付加価値を持つ場合もある．すなわち，以下のパラメータも考慮する必要がある．

(5) 情報使用時期

単に最近利用されたという新鮮度だけでなく，決まった時期に再利用されるといった周期性にも着目すべきである[3].

(6) 情報使用場所

(7) 情報使用者

利用個人だけでなく，利用者の役職，業務区分，人的交流などによる偏りにも着目する必要がある．

3.2 カテゴリ分類アルゴリズム

3.1 で挙げたパラメータのうち，大半は情報の属性，あるいはサービスの利用履歴などにより取得可能である．しかし，情報の内容にかかわる要素（カテゴリ）は抽出が困難である．その理由としては，

・扱う情報がもともと断片的であり，情報の分類基準の策定が困難である[4].

・情報の意味を代表するキーワードの精度よい取得が困難である．

・カテゴリの分類基準には，組織固有要素や時間変動要素も考えられ，固定的な分類基準を設けるのが難しい．

などが考えられる．

本検討では，組織固有要因や時間変動要因に対応して動的にカテゴリ分類を実現するため，情報中に共存する複数のキーワードの共起関係に着目し，同じ情報中に揃って出現する頻度の高いキーワードを同じカテゴリに属すると見なしたカテゴリ分類手法を検討する．具体的手法は以下に示す方法によりカテゴリ分類を試みた．

(1) 情報の文中より複数のキーワードを抽出する．

(2) 抽出したキーワードの関連性を出現頻度，出現位置によりスコア付けし，キーワードn, m 間の関連度（キーワード結合度）k(n, m)をスコアの比を用いて算出する．

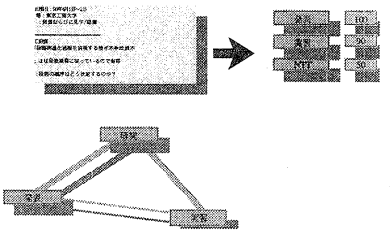

図1　キーワード結合度の算出

(3) 算出したキーワード結合度を情報数分加算したテーブル（キーワード結合度テーブル）を作成する．

(4)キーワード結合度テーブルの各要素をもとに，キーワードn, m 間の距離r(n,m)を計算する（キーワード組織化テーブル）．なお，距離の算出には以下の式を用いる．

$$1/r(n,m)=1/k(n,m)+\Sigma_{j\neq m,n}\{k(n,j)+k(j,m)/(k(n,j)*k(j,m))\}$$

(5) キーワード組織化テーブルをもとに，キーワード間の距離をリンクとして表示する．このとき，距離がある閾値以下になったものについてはリンクを削除する．

(6) 閾値を調整し，所望数のキーワード群に分割された時点で，それぞれのキーワードの組をカテゴリとする．但し，単独キーワードとして分割されたものについては，カテゴリとして扱わない．

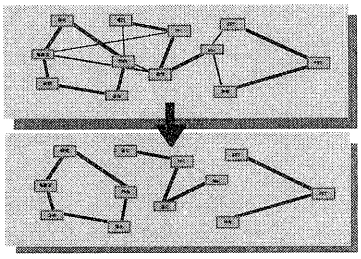

図2　カテゴリの分類

3.3　カテゴリ自動分類の評価

前節で提案したアルゴリズムの妥当性検証のため，実際の組織においてインフォーマルに扱われる情報を対象に自動分類を試みる．ここでは，２０名程度構成されたグループで使用されているネットニュースに蓄積された情報を対象とした．このグループではグループ内の雑多なインフォーマル情報の交換をローカルに構築したニュースグループを用いて行っており，年200件程度の投稿数がある．このうち，最近１年分の登録情報から任意に５０件を選び，前節のアルゴリズムによるカテゴリ分類を試みた．なお，各情報からは最大７つのキーワードを頻出順に抽出することとした．

妥当性分類の指標としては，意味的な妥当性（類似した意味の情報毎に分かれるか），分布的な妥当性（特定のカテゴリに偏らず分類そのものができるか）があるが，ここではまず後者に着目して検証する．

表1　カテゴリ分類結果（修正前）

研究，技術，（人名）	153
エージェント，音声，サブ	8
調査，送付，依頼	5
レク，委員，新年会	5
計算，プログラム，仮想	5

カテゴリ分類の結果を表１に示す．カテゴリ名は分類されたカテゴリの代表的なキーワード３種の組として示す．右側の数字はカテゴリ内のキーワードの個数を示す．また，（人名）はこのニュースグループに多く記事を投稿する利用者の姓である．計５つのカテゴリに分割されているが，大部分のキーワードは「研究，技術，（人名）」というカテゴリに偏っており，「分類」という観点からは本アルゴリズムが十分に機能しているとはいえない．

表2　他のキーワードとの関連の強さ

順位	キーワード	関連の強さ
1	研究	1625
2	（人名）	1250
3	・	815
4	システム	787
5	no	780

カテゴリ分割が十分に行われない理由として，複数のキーワードに対して強いリンクを持つ，すなわち距離 r(n, m) の値が小さいリンクを持つキーワードの存在が考えられる．表２に，他のキーワードとの関連性の強いキーワードの例を示す．表中の「関連の強さ」は，キーワード結合度 k(n, m) の総和を示す．この値が大きいということは，複数のキーワードに対して強いキーワード結合度k(n, m)を持つものであり，結果的に距離 r(n, m) の値が小さいリンクを多数持つキーワードと考えられる．表２では，「・」（中点）や「no」など，キーワードとして意味をなさないものが，情報文中では高頻度で使用されるために，高い「関連の強さ」を持っていることが，カテゴリ分類時の障害になっていると考えられる（図３）．また，ここで使用した情報はネットニュースのものであるため，（人名）についても，キーワードの抽出個所が主に冒頭部（「＊＊です」と名乗っている部分）や，末尾の署名として抽出されたものがほとんどである．これらの高頻出な不要キーワードの除去が本アルゴリズム上の課題である．

これらの不要キーワードが多く抽出されるのは，単語解析のアルゴリズムや使用辞書に大きく依存するため，完全に消し去ることは困難である．しかし，以下の手法により，ある程度は減らせることが期待できる．

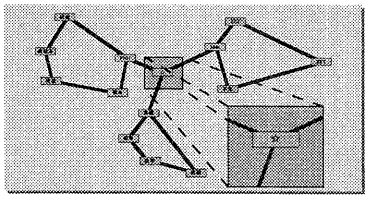

図3　不要語の存在によるカテゴリ分割失敗例

(1) ヒューリスティックの採用
　署名の削除，URL 等の削除，１文字キーワードの削除などの処理を加えることにより，明らかに不適切な個所からのキーワード抽出を抑制する．
(2) 不要語辞書／必要語辞書の導入
　冗長あるいは不適切な単語を予め不要語辞書として登録しておき，抽出時にこれに含まれる単語を削除する．逆に，優先的に採用する単語を必要語辞書として登録し，キーワードの種類が発散するのを防ぐ．
などの方法により，抽出キーワードの精度を上げることは可能である．
　以上の方法により，不要キーワード削除を行ったときのカテゴリ分類結果を表３に示す．

表3　カテゴリ分類結果（修正後）

研究，システム，端末	146
mail，ippan，alias	7
送付，調査，伝票	3
（人名２），hca，niftyserve	3
その他（分類先なし）	48

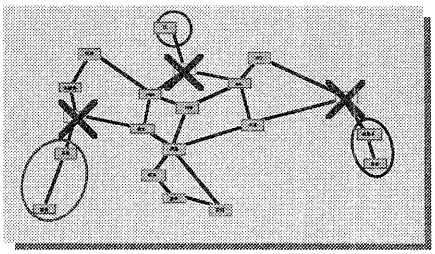

図4　不要語削除によるカテゴリ分類

表1同様に，特定のカテゴリにキーワードが集中する状態は続いているが，一方で，単独キーワードに分割され，「その他」扱いになったキーワードが増えている．不要語によって作られていたリンクを遮断することで，カテゴリの細分化は進んでいると考えられる．また，単独キーワードで分割されたため「その他」になったキーワードについても，登録情報の増加とともに，他のキーワードとの結びつきを強め，単独のカテゴリに成長する可能性を残している．

4. 情報組織化表現法の検討

4.1　情報組織化表現方法

　情報間の関連づけを行う方法としては，以下の２つが考えられる．
(1) 情報をある属性で分類し，情報群中でその情報の位置付けを提示する．
（例）・電子メール，ネットニュース（時間軸による分類）
　　・Yahoo 等の階層検索サービス（情報のカテゴリによる分類）
(2) ある情報に着目し，その情報に関連情報をリストアップする．
（例）・電子メール，ネットニュース：スレッドによる関連づけ
　　・FISH[4]：キーワードによる関連づけ

4.2.　情報組織化システム KnowHow Navigator の提案

　前節で述べた情報組織化表現方法を用いた情報組織化システム KnowHow Navigator の試作を行った．本システムは WWW 上にCGI ベースで構築され（図5），外部コマンドや CGI 中の複数の組織化モジュールを用いて情報の組織化を行い，そのパラメータをRDB に蓄積している．

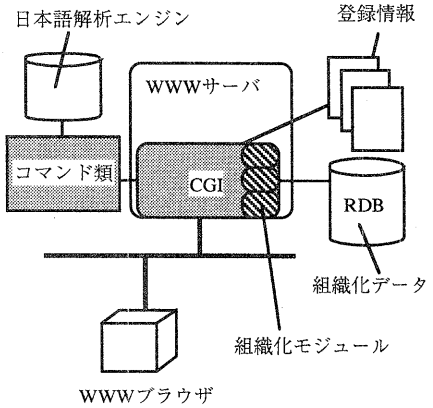

図5　システム構成

外部コマンドや組織化モジュールは，3.1で示した複数の組織化パラメータの抽出に対応し，追加削除が容易な設計となっている．登録情報は基本的に電子メールと同等のフォーマットを持つテキストファイルであり，CGI上からの登録の他，電子メールやネットニュース上からのコンバートも容易に行なえるため，情報登録の負荷も合わせて低減できる．

KnowHow Navigatorの組織化機能は2つあり，4.1で示した2つのアプローチにそれぞれ対応した機能を持たせている．

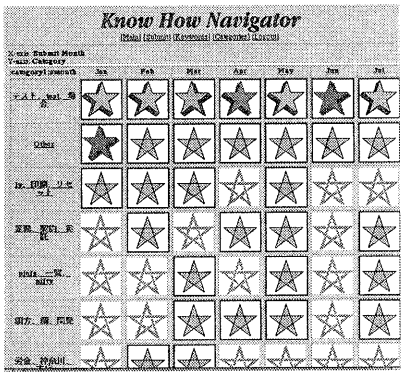

図6　情報組織化画面例

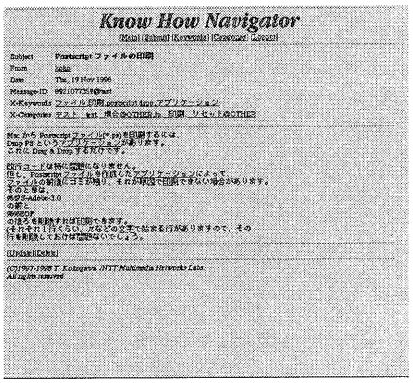

図7　情報リンク画面例

(1) 情報群の分布を可視化する機能（図6）

システムに組み込んだ組織化モジュールより任意に2軸を選び，情報分布を2次元表示する．情報数の大小はアイコンの模様（差し換えは可能）で判別できるようにしている．さらに，軸をクリックすることで，サブカテゴリがある場合はその表示も可能．

(2) 情報間の関連を可視化する機能（図7）

当研究グループにて過去に開発したFISH[2]の機能を応用し，関連キーワードを用いたリンクを情報中に埋め込むとともに，登録者，カテゴリなどに対しても関連リンクを作成する．

5. まとめ

本検討では，組織におけるインフォーマル情報の共有における課題として，情報組織化の要求条件を整理し，支援アプローチの検討を行った．情報中に現れるキーワード間の関係を用いたカテゴリ自動分類のアルゴリズムを提案し，実装を行った．その結果，不要キーワードの混在によるカテゴリ分類処理の問題が生じやすく，キーワードの適切な抽出が分類アルゴリズムの有効性とともに大きな問題となることが示された．

一方で，情報組織化にはカテゴリ分類だけでなく，情報の生成・消費に拘わるさまざまな状況（時間，人，場所など）も考慮する必要があり，これらの情報を利用者の視点に応じて適用した情報組織化により補完できると考えられる．この観点で情報組織化システムKnowHow Navigatorを実装した．今後フィールド運用を通じて評価していく予定である．

参考文献

[1] Conklin, E. J., Capturing Organizational Memory", *In Groupware '92*, pp.133-137, Morgan Kaufmann Publishers, 1992.

[2] 爰川, "インフォーマル情報の共有支援に向けて", DiCoMoワークショップ, 107, 1997.

[3] 門脇ほか, 情報取得アウェアネスによる組織情報の共有促進支援, 人工知能学会誌, 投稿中。

[4] 稲田ほか, "マルチメディア情報共有システムの検討", 信学技報, 1996.

[5] Seki, Y., Yamakami, T., and Shimizu, A., "Flexible Information Sharing and Handling system : Towards Knowledge Propagation," IEICE Trans. Commun., March, 1994.