

FISH の利用履歴分析に基づくグループ情報共有支援の検討

爰川知宏†, 荒金 陽助††, 杉田恵三†

†NTT 通信網研究所
〒238-03 神奈川県横須賀市武 1-2356-523A
††東京工業大学 総合理工学研究科
〒226 神奈川県横浜市緑区長津田町 4259

あらし グループ内での情報共有行動は、ユーザのサービス利用に対する意識や、グループ内でのサービスに対する共通認識などに強く影響される。本検討では、ノウハウ蓄積システム FISH の利用動向の解析により、情報検索行動の失敗に着目して解析を行うことで、グループ情報共有の本質的課題を明らかにし、その支援方式の検討を行う。ユーザのサービス利用形態により情報検索パターンを4つに分類し、それぞれのパターンにおける支援方式について検討を行った。

和文キーワード 情報検索、グループウェア、アウェアネス

FISH Usage Analysis to Support Group Information Sharing

Tomohiro KOKOGAWA†, Yosuke ARAGANE††, and Keizo SUGITA†
†NTT Telecommunication Networks Laboratories
e-mail: {koko, sugita}@nttmhs.ntt.jp
1-2356-523A, Take, Yokosuka, Kanagawa, 238-03 JAPAN
††Tokyo Institute of Technology
4259, Nagatsuda-cho, Midori-ku, Yokohama, Kanagawa 226 JAPAN
e-mail: yaragane@titech.ac.jp

Abstract The group communication is depend on the recognition of service usage of each user or whole group. By analysing FISH (Flexible Information Sharing and Handling system) usage pattern, especially, the missing pattern of information retrieval from FISH, we discussed the essential problem of group information sharing, and suggested their support method. By analysing FISH usage of each user, and classifying the information retrieval pattern by four types, information sharing support methods are suggested for each user types.

Keywords Information retrieval, Groupware, Awareness

1. はじめに

World-Wide Web (WWW) に代表されるような、Internet 上で情報を発信/収集するための環境の急速な普及や、Lotus Notes 等の組織内情報共有を支援する市販グループウェアアプリケーションの普及により、オフィスをとりにまく情報環境は急速に変化しつつある。そのような状況の中で、数多くの情報源から情報洪水を避け、必要な情報を効率的に収集するための情報検索に関する技術がますます重要になってくる。

情報検索に関しては、これまでも意味検索や全文検索の高速化など、数多くのアプローチがなされており、Yahoo のような WWW 上の情報検索サービスをはじめ、様々なサービスやアプリケーションへの実装も多く行われている。

一方、オフィスで扱われる情報には、定型化可能な公式文書やマニュアルだけでなく、業務を進める上でグループの各人が培ってきた種々のノウハウと呼ばれる情報もある。このような情報は動的、断片的で、定型化して扱うことが困難である。検索する側にとっても、断片的な情報であるが故、そのノウハウ情報を得るための要求を明確化することが難しい。

このような動的、断片的、非定型なノウハウ情報の蓄積/共有を目的として、ノウハウ蓄積システム FISH [Seki 94] がある。このシステムでは、情報登録ユーザが任意に付与したキーワードを検索プロファイルとして、散在する情報間でリンクを張ることで、ユーザがたどり着いた情報から、その情報に関連する他のノウハウ情報へ容易にアクセスすることができる。しかし、最初にアクセスするときに、利用者の想定したキーワードが、システムが許容するキーワードと一致しないために、検索の失敗が生ずる場合がある。

本検討では、情報検索の課題のうち、情報検索の失敗に着眼した検討を行う。情報検索の失敗には、

(1) 広い概念を表わすプロファイルで検索したためヒットする件数が多すぎて、必要な情報が絞り込めない、(2) プロファイルの認識のずれで、ヒットした情報に必要な情報が含まれていない、(3) システムが許容しないプロファイルを選んだため、情報がヒットしない、が考えられる。何れの場合も、そのシステムにおいて有効と考えられる検索プロファイルに対するユーザの意識のずれや、過去の利用実績、グループ内部での意識の共有などが作用している。これらは情報検索において本質的な課題であり、単純な検索効率や検索速度だけでなく、ユーザの行動そのものにも着目した情報検索に関する議論が行われなければならない。本検討では、情報検索の失敗パターンの解析、分類を行うことで、グループ内の情報共有行動における本質的な課題を明らかにし、その支援方式について検討する。

2. 情報検索失敗パターンの分類

情報検索とは、ユーザが必要とする情報に対して漠然と抱いたイメージから検索プロファイルを作成し、それを介してサービス上で情報のフィルタリングを行うことである。その検索プロファイルをサービスが受け付けないとき、あるいは検索プロファイルがユーザの当初の要求を十分に表していないとき、情報検索に失敗する。

キーワードをプロファイルとして情報を検索するシステムの場合、キーワードの誤りには以下のケースが考えられる。

- (a) 単純なスペルミス
- (b) 大文字/小文字、全角/半角など
(システム上の制約)
- (c) 意味的に類似しているが、サービス上には存在しなかったキーワード
- (d) 操作上のミス
(コマンド使用方法の間違い)

また、検索失敗のパターンには以下のような特徴が考えられる。

- ・失敗の連続性
情報検索に失敗したとき、成功するまで試行錯誤的にプロファイルを修正して情報検索を繰り返す。
- ・失敗の反復性
一度検索に失敗したプロファイルについても、その検索プロファイルに対する思い込みが強い場合、再び同じ失敗を繰り返す。
- ・失敗の共有
他人と同じプロファイルで失敗する場合がある。システムを導入しているグループ内で、そのプロファイルが共有されているが、システムには反映されていない。

以下では、それぞれのケースについて、具体的なシステムの利用動向観察により解析を行う。

3. 情報検索の解析

ここでは対象サービスとして、GoldFISH [関 95] の利用動向を用いる。GoldFISH は、先に述べたノウハウ蓄積システム FISH を検索処理能力を向上させた上で分散サーバ対応に拡張したものであり、情報の管理方式は FISH と同様である。サービスの利用はコマンドインタフェースを用い、`fpick [keyword]` で情報検索、`fshow [number]` で情報参照、`fcomp` コマンドで情報登録が行える。

対象グループは 15 名前後の構成員からなる研究開発業務を行う組織であり、GoldFISH を導入してから約 2 年、先代の FISH からは約 5 年の使用実績があり、合わせて約 2,000 件の登録情報をもつ。このシステムにおいては、情報登録ユーザは任意に情報のキーワードを付与しており、強制を伴う意思統一は行っていない。また、実際に登録されている情報

は、時刻表や電話番号のようなテーブルデータ、伝票処理などの処理手順や注意点を記した業務ノウハウ、研究業務に関する情報など、広く業務全般にわたる情報が登録されている。さらに、どのキーワードで検索されたか、どの情報を参照したか、などの利用履歴も継続的に収集しており、このうち本検討では、95年2月より7月までの半年間の利用履歴データを対象として解析を行う。

3.1 検索失敗の連続性

ユーザそれぞれに対し、検索失敗をした後の5入力までにまた検索失敗をすることを再NGと名付け、トータルのNG(失敗)回数に対する割合を調べた結果を表1に示す。

表1 NGと再NGの関係

月	NG数	再NG数	NG確率
2	17	9	0.5294
3	70	44	0.6286
4	36	22	0.6111
5	44	29	0.651
6	90	70	0.7778
7	76	58	0.76

この結果より、検索失敗後5入力以内で再度検索に失敗する確率は各月とも約6割程度と高い。すなわち、検索に失敗したことで6割程度は検索成功まで試行錯誤を行っていると考えることができる。

3.2 検索失敗キーワードの解析

以下のような機能をもつプログラム(dpick)を作成し、利用履歴を入力したときのデータを解析した。このプログラムは解析だけでなく、実際のGoldFISH自体のコマンドを置き換えてキーワード入力の手助けを行うプログラムとして利用することも可能である。

- ・GoldFISHであるキーワードAにより検索した結果が失敗である場合に、サブキーワード辞書に対する登録が発生する。
- ・同一ユーザがその後初めてある他のキーワードBで検索に成功した場合、キーワードBをキーワードAのサブキーワードとしてサブキーワード辞書に登録する。
- ・次に語Aが検索された場合、サブキーワード辞書はキーワードBを候補語として表示する。
- ・新たにキーワードAを持つカードが登録された場合には、Aの候補語は表示されなくなる。

このプログラムに半年分のGoldFISH上での情報検索履歴を投入した結果を表2に示す。

表2 検索履歴入力結果

月	検索数	NG数	ヒット数	ヒット率	辞書登録	新規登録
2	161	15	0	0	14	14
3	510	68	15	22.1	65	51
4	454	50	7	14	108	43
5	364	41	9	22	140	32
6	589	101	32	31.7	201	61
7	457	63	21	33.3	243	42

上記プログラムを実行することで、検索に失敗したキーワードは成功したキーワードのサブキーワードとして関連づけられる。半年間で蓄積された検索失敗のキーワードは243語それぞれに対して生成したサブキーワードは274語であり、1つのキーワードに対してのサブキーワードの数は16個であった。その内訳は1つのキーワードに対して生成されるサブキーワード数は、以下に示す結果となった。

- ・1つのサブキーワード：210 キーワード
- ・2つのサブキーワード：18 キーワード
- ・3つのサブキーワード：5 キーワード
- ・4つのサブキーワード：3 キーワード
- ・5つのサブキーワード：0 キーワード
- ・6つのサブキーワード：1 キーワード

さらに、親キーワード(正解キーワード)とサブキーワードの関連を分類した結果は以下のとおり。

- (a) 関連はあるが存在しないキーワード
 - 1) 関連なし/不明
(途中で断念?失敗の経緯不明)
 - 2) 直後の成功と関連あり
(意味的に類似しており、試行錯誤で正解にたどりついたものと考えられる)
- (b) 単純タイプミス (tel -> tek など)
- (c) 大文字/小文字 (Mac -> mac など)
- (d) 操作ミス (fshow 1 -> fpick 1 など、類似したコマンドの誤起動)

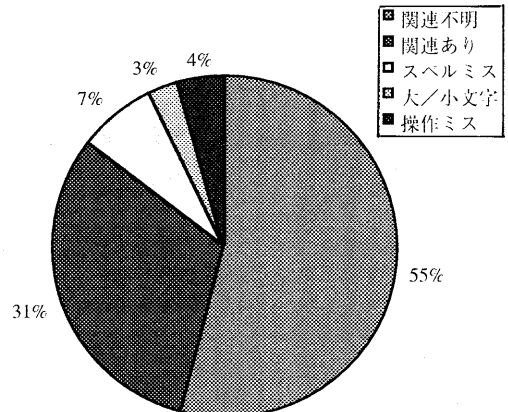


図1 検索失敗キーワードの分類

この結果から、検索失敗のうち、約半数は最終的に目的の検索プロファイルにたどり着くことができたと考えられる。

3. 3 検索の反復性、共通性

同じ失敗を繰り返しているなら、上記履歴辞書に対するヒット率が高いことが予想される。また、失敗が他のメンバと共有されているなら、他人の作った履歴辞書に対するヒット率が高くなることが予想される。ここで、自己ヒット率と相互ヒット率という概念を導入する。自己ヒットとは、検索失敗と検索成功の組合せより自分で辞書に登録したキーワードに対して、後に自分自身でまたそのキーワードを検索して辞書にヒットすることである。また逆に相互ヒットとは、他人が辞書に登録したキーワードについて検索し、辞書にヒットすることである。自己ヒット率、相互ヒット率はそれぞれのヒット数その月の検索失敗回数(NG数)で割ることによって、ヒット数を正規化したものである。

x 月の自己(相互)ヒット率 = x 月の自己(相互)ヒット数 / x 月の検索失敗回数(NG数)

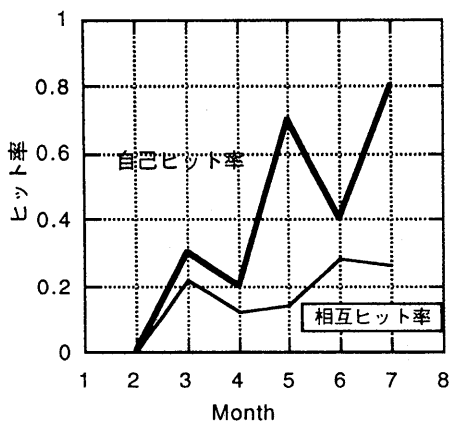


図2 自己ヒット率と相互ヒット率の関連

図2に自己ヒットと相互ヒットの変遷を示す。どちらのヒット率も、時間の経過にともなう履歴辞書の充実により漸増しているが、自己ヒットの伸びが比較的大きい。さらに、自己ヒット数とNG数の間には線形相関が予想される。一方で、相互ヒット数に関しては、同様の相関は見い出せなかった。自己ヒットはその月の中である程度完結しており、遠くさかのほった月に、自分が辞書に登録したキーワードを参照することは、つい最近に自分が登録したキーワードを参照するのに比べて極端に少ないことが予想される。したがって、短い期間の中で同じキーワードをたびたび検索し、失敗しているものと考えられる。他方、相互ヒット数はその月の検索失

敗数に関係なく漸増していること、またその率(=数)が少ないことから、辞書に登録された時期にかかわらず検索されその検索パターンは単発的であることが予想される。

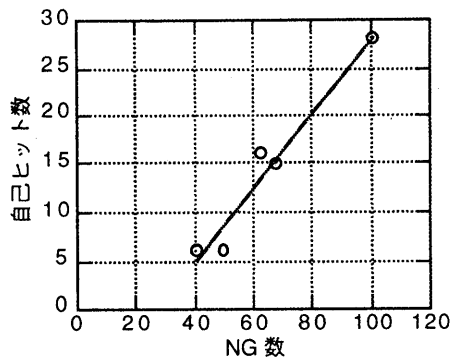


図3 NG と自己ヒット数の関係

4. ユーザ毎の傾向

対象ユーザ 14 名全体の推移は以上のようなものであるが、ここでは、各ユーザごとに焦点を当てて議論する。GoldFISH の利用状況に応じて4つのタイプに分類し考察する。

- ・type-I:月ごとの最大検索数が15を複数回越えないもの。(5名)
- ・type-II:月ごとの最大検索数が15を複数回越えるが、検索数10未満の月があるもの。(3名)
- ・type-III:type-I,II以外のもので、検索数の月平均が50未満のもの。(2名)
- ・type-IV:type-I,II以外のもので、検索数の月平均が50以上のもの。(3名)

◆type-I:ほとんど利用しないユーザ

構成員は5名とすべてのタイプ中最大であるが、失敗する回数は検索回数に比例して極端に低く、月平均にして1回未満である。ログデータの所見は同一のキーワードを検索する例が非常に多く、失敗しても簡単なスペルミスが原因で次の入力時に改正されているか、失敗の後に思考錯誤することはなくすぐにあきらめてしまうといった特徴があげられる。このタイプのユーザは、他の手段によるノウハウの入手を行っている等の理由で、「GoldFISH」に対する認識がデータベースや電子掲示版等の静的なデータの蓄積というものであると考えられる。

◆type-II:type-Iよりは利用するが月ごとのばらつきがあるユーザ

このタイプのユーザは失敗する回数が月平均で1.8~3.0と比較的大きく、また、試行錯誤により失敗の連続性が強くが見られるといった特徴を有する。

この試行錯誤ということからも、「GoldFISH」をノウハウ蓄積システムとしてとらえ、動的なノウハウを検索しているユーザでとえられる。

type-IIのユーザのサブキーワード辞書のヒット率について考える。ヒット数は月あたり0~3と失敗数が少ない分小さい値である。それだけに、一度サブキーワード辞書がヒットすれば検索失敗数に対するヒット率は高くなる。

ここでどのようなキーワードで検索に失敗し、さらにヒットするのかがログデータを見てみると2つのことが解る。

- ・同意語が多数見つかる。
- ・相互ヒット率が、比較的高い。

1.より同意語が多数見つかり、しかも時期的に連続することから、このユーザが検索失敗した直後に試行錯誤を行っていると考えられる。したがって、履歴辞書が充実すれば1回目の検索失敗時に正しいと思われる検索プロファイルを示すことができる。また、相互ヒット率が高いことから、他のユーザとの情報共有について積極的であり、履歴辞書の充実による相互ヒット率の向上によって履歴辞書ヒット率の増加が多く見込まれるユーザである。

◆type-III:比較的恒常的に利用するユーザ

type-III,IVは共に「GoldFISH」を動的ノウハウ蓄積システムと位置付け、使いこなしているユーザであると考えられる。図4はこのタイプに属するあるユーザの検索失敗数とヒット率との関係を示したものである。図4より、検索失敗数が増加した月にヒット率も増加するという関係が得られる。以上より、このタイプのユーザは短期間に検索失敗をおこす同一のキーワードを複数回使用していると考えられる。したがって自分自身で辞書に登録したキーワードを短期間に何度も検索していることで、自己ヒット率が高いことが予想できる。

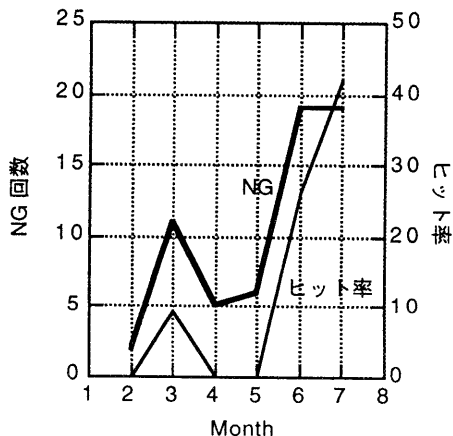


図4 type-III ユーザのNG回数とヒット率の関係

◆type-IV:恒常的に利用するユーザ

回数的には最も「GoldFISH」を利用していると考えられるユーザである。このタイプのユーザの特徴を、入力数に対する検索失敗数とヒット数の割合から検証する。

表3より、ヒット数の比がtype-IIIに比べて少なく、入力数の増加がそのままヒット数(NG数)の増加には結び付かないことがわかる。思うにこのタイプのユーザは「GoldFISH」をノウハウの詰まった手帳のようなものと、とらえていると考えられる。ノウハウ自体は忘れてしまっても、それを導き出すキーワードは記憶にあり、それを使って必要なノウハウを必要な時に頻りに引き出していると推測される。

表3 検索失敗とヒット率

user	user31	user32	user41	user42	user43
検索失敗率	28.1	13.3	7.65	7.8	22.7
ヒット率	6.33	7.52	1.02	2.06	5.11

以上それぞれのタイプのユーザの特徴をまとめると以下のようになる。

表4 ユーザタイプ毎の特徴の分類

タイプ	I	II	III	IV
利用頻度	小←	ばらつき	やや大	→大
失敗頻度	小←	大	大	→小
参照キーワード	限定←	不特定	不特定	→個別限定
失敗の連続	小←	大	やや大	→小
失敗の反復	小←	やや大	大	→小
失敗の共通	小←	大	やや大	→小

5. 検索支援方式の検討

4節で得られたユーザタイプ毎の特徴より、情報検索支援方式について検討する。

type-I ユーザの場合、利用頻度が少なく、検索失敗の頻度も少ないが、既知の限定されたカードしか見ない傾向が強く、新しい情報に気づきにくいと考えられる。したがって、検索プロファイルに対するアウェアネスを与えるような支援、関連語へのリンク機能の強化で、検索プロファイルを意識する必要なく周囲の関連情報に対するアウェアネスを与えるアプローチが必要である。

試行錯誤が多い type-II ユーザの場合、スペルミスや操作ミスのような単純なミスも含め、正しい検索プロファイルに導くような支援が必要である。支援方式としては、意味検索などの曖昧性をもたせた検索プロファイルを用いてヒット率を上げることも考えられるが、2節の解析で用いた、検索失敗の履歴辞書を用いたアプローチにおいても、図1より、半数程度の検索失敗が原因が明確であることから、支援効果が大きいことが予想される。加えて、type-III ユーザにおいては自己ヒット率が高い、すなわ

ち、個人の検索プロフィールにおいて、検索失敗の反復性が強い場合、履歴辞書の効用は期待できる。

type-IV のヘビーユーザの場合は、失敗は少ないが、検索対象が自分自身が登録した個人情報に偏る傾向が考えられるので、関連情報へのリンク機能の強化など type-I と同様の、他の情報へ気付かせるためのアプローチが必要と考えられる。

本検討で解析のために用いた履歴辞書を用いた検索支援の効果は、ヒット率が3割程度であるが、頭打ちの傾向が現われている(図5)。これは、履歴辞書の充実の一方で、新規情報が次々と登録され、それによって検索プロフィールもさらに多様化するためと考えられる。

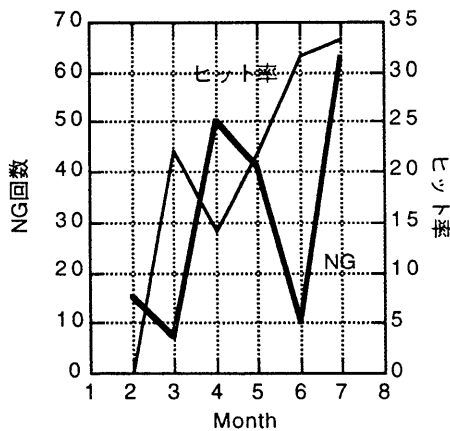


図5 履歴辞書の導入効果

6. まとめ

本検討では、GoldFISH の利用履歴をもとに、情報検索の失敗パターンに着目した解析を行った。この解析により、検索失敗行動の共通パターン(連続性、反復性、共通性)に関するグループの特徴が得られた。さらに、ユーザタイプ毎の分類により、それぞれのパターンの差異を明らかにし、支援のための指針を与えることを可能にした。特に、サービスの定着に強い影響を与えられると考えられる、自ら試行錯誤を繰り返して必要な情報を集める type-II, III のユーザの支援に、過去の検索履歴を用いて検索のヒット率を上げるようなアプローチが有効であると考えられる。

以上の検討より、グループ情報共有における情報検索は、ユーザの行動パターンの影響を強く受けており、検索速度や検索効率だけにとらわれず、支援を検討する必要がある。WWW の利用者の情報収集行動で Social HyperText が形成されていくということも報告されており[Erickson 96]、ユーザやその所属するグループの社会的な行動プロセスに着目した支

援方式の検討が今後必要である。

謝辞

本研究は 1995 年 8 月の NTT 通信網研究所における学外実習プログラムの一環で行われました。本研究の機会を与えてくださった、木下研作ネットワークインテグレーション研究部長に感謝します。

参考文献

- [Erickson 96] Thomas Erickson, "The World-Wide Web as Social Hypertext", CACM, Vol. 39, No. 1, pp. 15-17, 1996.
- [Seki 94] Seki, Y., Yamakami, T., and Shimizu, A., "Flexible Information Sharing and Handling system: Towards Knowledge Propagation", IEICE Trans. Commun., March 1994.
- [関 95] 関 良明, "分散型ノウハウ蓄積システム GoldFISH における分散環境への適応", 情処論, Vol. 36, No. 6, pp. 1359-1366, June 1995.