

複数人の経験共有による情報検索手法

板野 愛[†] 小林 一郎[‡]

^{†‡}お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻

E-mail: [†] {itano.ai, koba}@is.ocha.ac.jp

1 はじめに

近年、様々な検索ツールの利用が可能になっているが、依然、キーワードに依存した検索が主流を占めている。このことから、人の経験に基づき時間や記憶をきっかけにした検索システムの開発が行われている。それらの研究の多くはユーザ個人の経験のみに基づいた検索手法となっている。これに対して本研究では、家族などの小さな共同体において複数人によって情報が自然に共有される場合に、複数人のイベントを介在して個人の情報を効率的に検索する手法を提案し、その手法に基づくシステムを開発することを目的とする。

2 経験共有に基づく検索

2.1 経験共有に基づくイベント増加

複数人で経験を共有することにより、個人のイベントだけでなく、自分が知っているコミュニティ内の他人のイベントを通して自分のファイルを検索することが出来る。複数人のイベントを登録することで、自分だけのイベントを検索する場合に比べ、検索の再現率を上げることができる。

2.2 イベントとデータの関連性

イベントと取得されたデータには関連性が存在し、特定のイベントごとに特定のファイルが関連付けられるという、イベントのタイプとファイルの種類にある程度の共起関係が存在する。イベントとファイルの対応関係を明確にすることによりイベントを機軸にした特定ファイルの検索精度向上を検討する。

表 1 イベントとデータの関連の例

イベント	ファイル
学校	Microsoft Office
仕事	Microsoft Office, PDF
旅行	画像・動画

2.3 イベントを基軸とした情報検索

イベントを機軸とした情報検索は、イベントを考慮して、検索対象ファイルの絞り込みを行う操作に相当する。以下に、イベントに対してファイルを特定する方策を示す。

(i) 時間的一致性

ファイルの生起がイベントの生起と同時刻の場合、時間的一致性を捉えることにより対象ファイルの絞り込みを行う。

(ii) イベントの前後における時間的一致性

ファイルの生起がイベントの生起の前後の時間帯である場合、その時間帯における時間的一致性を捉えることにより対象ファイルの絞り込みを行う。また、本研究ではイベントの前後の他に「最近」や「結構前」という言葉を使用している。「最近」とは当日からの日付の差が大きければ大きい程該当する確率が少なくなり、「結構前」では、ファイルが一週間以前に生成された場合、当日からの差が大きい程確率を増やすものとした。

(iii) 自分と他人のイベントとの相対関係

上記(i)～(ii)におけるイベントを利用したファイル特定の手段は、個人のイベントだけでなく他人のイベントに関しても適用可能である。これらに加え、個人のイベントと他人のイベントの生起の相対関係により対象とするファイルを絞り込むことができる。

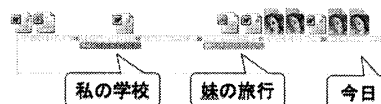


図 1 ファイルと家族のスケジュール

3 検索結果の評価方法

与えられる条件は「時」「ファイルタイプ」「所有者とイベント」である。「所有者とイベント」は 0～3 個指定できる。まず、チェックした確実な条件以外に該当するファイルを検索対象から外し、検索対象となるファイルに値をつける。一致度は、それぞれの項目の一致度の総和からそれぞれの完全一致度の総和を割ったものとしている。

4 提案する検索手法の検証

上に提案した情報検索手法を検証する。本システムでは、一般的な検索システムと比較して、ユーザが求めるファイルに対してどれだけ適した情報を提供できるかを評価する。検索対象となるファイルは PC で収集した実際の 1 ヶ月分のファイル 71 件である。ファイルのタイプはワードやエクセル、画像など様々なものが含まれる。これらを検索対象とし自分のスケジュールと家族のスケジュールを用いて検索を行う。

4.1 検索条件

ユーザは以下の検索条件を指定できる。

- ① ファイルを作成した大まかな日時
- ② ファイルタイプ
- ③ 複数人のイベントとの同期

更に、①～③における検索条件を一つ選び基軸として検索することができる。

An Information Retrieval Method Based on Sharing Experiences in a Small Community

[†]Ai ITANO [‡]Ichiro KOBAYASHI

Graduate School of Humanities and Sciences, Advanced Sciences, Kobayashi Laboratory Ochanomizu University

E-mail: ^{†‡} {ai_i, koba}@is.ocha.ac.jp

4.2 検索結果と関連候補ファイルの提示

検索結果は、指定された条件にマッチした検索条件が評価の高い順に表示される。評価？また、基軸とした検索条件のみに関連するファイルを 1 つ提示することになっている。これは、ユーザが指定した確かな情報のみを条件として検索した結果を提示している。

4.3 検索結果の評価

実際に PC に保存された 1 ヶ月分のファイル 71 件を検索対象とする。また、自分や家族のイベントの組み合わせは 20 件となった。その中で「中間発表.ppt」というファイルを探したい場合に、(1)～(3)の条件の下、検索を行った。

(1) 検索条件に時間とファイルタイプを選択した場合

「最近使用した、拡張子が不明なファイル」を指定した場合、結果として 71 件あるファイルから「最近」という時間情報に適合した 30 件が検索結果として反映された。最上位に出た結果は「010.jpg」であった。検索結果 30 件の中に中間発表.ppt というファイルは存在するが、結果、上位には表示されていなかった。



図 2 時間が指定されている場合

(2) 検索条件の(1)に更に自分のイベントや他人のイベントを指定した場合

「最近使用した、拡張子が不明なファイルで、私が学校に行っていて、妹が旅行に行っている時」と指定すると、検索結果は 8 件あり、最上位に出た結果は「中間発表メモ.txt」であった。

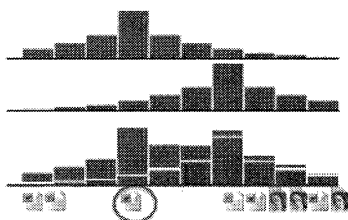


図 3 イベントが指定されている場合

(3) 検索条件の(2)に確実な情報のカテゴリを指定した場合

条件「最近使用したが、拡張子が不明なファイルで、私が学校に行っていて、妹が旅行に行っている時」を指定し、更に「最近」使用したのが確からしいとした場合、図 2 のような結果が得られた。検索結果は 8 件で、最上位にあるファイルは「中間発表.ppt」である。

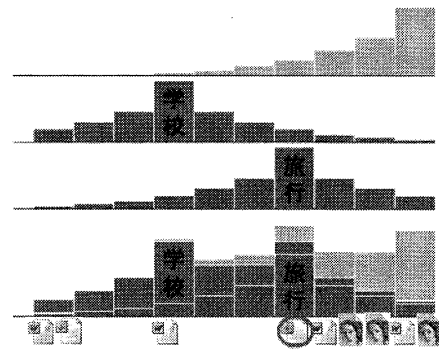


図 3 全ての条件を指定した場合

5 考察

提案した検索手法を利用し検索を行ったところ、曖昧な記憶であっても、複数の条件を指定することにより理想とするファイルを検索できることがわかった。3.2 検索結果の評価の(1)(2)(3)において、一般的な検索条件である(1)の条件では、検索結果が多く、絞り込みは難しいことがわかる。(2)の自分のイベントだけでなく他人のイベントを指定する場合、理想的な絞り込みは行えるものの、欲しいファイルは上位に表示されていなかった。

個人のイベントを中心に情報検索を行う際に、検索対象となるファイルの発生と個人のイベントが時間的に一致している際には、一般的にイベントを通じて対象ファイルを特定しやすい。また共同体のイベントが登録されている場合には、個人のイベントが登録されていた際の検索手法を含み、さらに、コミュニティ構成員のイベントを通じて対象となるファイルの検索が行えるようになることから検索の再現率が向上し、さらに構成員の複数のイベントに基づくファイルの絞り込みが行えるようになることから適合率も上げることが可能となることが検証された。

6 おわりに

本研究では、人が記憶しているイベントを通して、そのイベントに関連するファイルを検索するシステムを構築し、有効性を確かめた。検索結果とユーザ自身の理想順位とを比較して、有効であることを確認し、過去のイベントという知識の共有が自分のファイルを検索する際に役立つと考えられる。今後の課題として、イベントがない時間帯をファイルと結びつけること、更にどんなユーザにも利用可能なように拡張すること、が挙げられる。

参考文献

- [1] Jun Rekimoto, "Time-Machine Computing: A Time-centric Approach for the Information Environment", ACM UIST'99, 1999.
- [2] S. Dumais, E. Cutrell, J. Cadiz, G. Janck, R. Strin and D. C. Robbins: "Stuff I've Seen: A System for Personal Information Retrieval and Re-Use", Proc of SIGIR'08, pp. 72-79, 2003.
- [3] 松原靖子, 小林一郎 "ユーザのスケジュール情報と PC 内データ間の相関関係を利用したデスクトップ検索システムの開発" 電子情報通信学会技術研究報告. DE, データ工学, Vol.107, No.131 (20070625) pp. 31-36, 2007.