

# 「タグ付けで ノイズの少ない検索」

ただただし (tDiary.org)  
sho@spc.gr.jp

Google Desktop や MacOS X の Spotlight など、デスクトップ向けの検索ソリューションが一般的になってきた。ディスク容量の飛躍的な増大、それに伴うメールを主とする電子文書の扱いが増えていることが背景にあるのだろう。個人の電子データもすでに人力で分類・整理するのは無理がきていて、検索機能を使って探し出す方向にパラダイムを変えなければいけない時代になったということである。

情報を見つけ出すのに検索が有効なことは、Google や Yahoo! のような検索サイトを使って日常的にインターネット上の資源を探し出していればすぐに想像がつく。しかしながら、デスクトップ向けの検索機能は、これらの検索サイトと違い、ノイズが多くて目的のデータにたどり着くまでに時間がかかることが多い。検索サイトがリンク情報等を利用して重要度のランク付けを行っているのに対し、個人データを対象とした検索機能にはランク付けの手がかりがほとんどないためだ。そのため、検索結果を絞り込んでノイズを減らすために、なんらかの補助的な手法が必要になる。

## ソーシャルブックマーク

この絞り込みにはいくつかの手法が考えられるが、筆者が情報整理に好んで使っているのは、タグングである。タグングの考え方はテキストに限らないファイルへの情報付加のために生まれたが（MP3 ファイルの ID3 タグなど）、固定フィールドだけでなく、任意の情報を複数付加できるのがポイントである。

このタグングの考え方が広く知られるようになったのはソーシャルブックマークによってだと思うが、筆者も主にネット上のリソースをメモするために利用している。del.icio.us やはてなブックマークが著名だが、筆者が愛用しているのは入力情報が豊かな MM/Memo (図-1)

以下のURLに関するメモ  
[MM]トップページ - memo (もっと詳しく)  
http://1470.net/mm/

タイトル:

※省略時は上に表示されているURL/メディアのタイトルが使われます。

日付: 2006 / 3 / 25

メモ種別:  ←

評価:  未指定  1  2  3  4  5

ジャンル:  ←

※半角「|」で区切ると複数のジャンルを指定できます

自分がつけたジャンル: blog Google Security あとで Ajax OSS javascript Usability wiki Windows sbm 著作権 Apple Sony Spam search TrackBack はやぶさ book SR css livedoor shop palm

他の人がつけたジャンル: MM 更新情報 仕様変更 機能追加 JAVA folksonomy 技

図-1 MM/Memoの入力フォーム

である。

Web ブラウザでメモしておきたいページを見つけた時には、専用のブックマークレットを使ってそのページの URL やタイトルなどの情報を MM/Memo の入力フォームに自動的に入力し、必要とあればコメントをつけた上で、タグ（MM/Memo 上では「ジャンル」）を複数指定して保存する。タグは将来この情報を検索するための重要な手がかりになるので、十分な注意を払って指定する必要がある。筆者が心がけているのは、「未来の自分が検索時に使いそうなキーワードを付ける」という点だ。

通常、自分の情報を検索するという行為は、過去の自分が書いたことを思い出し、それをキーにするものだが、タグングは未来を想像するという逆の行為になる。過去の自分は最も信じられないものの1つなので、この方向性は正しいと筆者は信じる（未来の自分はその次に信じられないが）。





図-2 Picasa



図-3 iTunes

### 画像ファイルと紙文書の管理

ローカル資源の整理には、MacOS X を使っていれば Spotlight を使ってシステム全体でタギングを利用できる。Windows でも次世代ファイルシステム WinFS で属性の付加が可能になるはずだったが、Windows Vista には搭載されないようなので、筆者のように Windows を使わざるを得ない場合にはアプリケーションレベルで実現するしかない。

タギングが特に有効な場面として、全文検索が不可能なメディア系のファイル整理があげられる。筆者は主に、デジタルカメラで撮った写真と、紙文書の整理にタギングを活用している。

画像ファイルの整理に使っているのは Google が無償で配布している Picasa (図-2) だ。Picasa は高速スクロールを売りにした画像管理ソフトだが、管理手法にフォルダ分けを使わず、検索をメインに据えているのが特徴的である。検索対象は Exif データやファイル名であるが、そのほかに任意のタグ付けが可能になっている (Google 用語では「ラベル」)。

写真に限らず、フォルダによる階層構造管理は簡単に破綻する。たとえば家族旅行の写真は、旅行フォルダに入れるべきだろうか、家族のフォルダに入れるべきだろうか。両者がツリーの別々の枝にある場合には分類が不可能である。タグを使う場合、このような場合には「旅行」と「家族」という2つのラベルを付加することで解決できる。

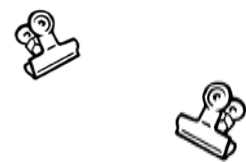
Picasa 上でタグ付けされた写真を探すときには、まず検索フィールドにタグを入力し、その上で日付などを元にした絞り込みを行う。あらかじめタグによって絞り込

まれているので、目的の写真を選び出すのは容易である。

紙の文書も、検索対象にするのが困難なメディアである。筆者はできる限りスキャナで取り込んで、PDF にして管理するようにしている。最近のスキャナは PDF 化の際に OCR を通してテキストレイヤを作成してくれるようなものもあるが、誤認識が少なくないことと、全文検索はノイズが多いことから、画像ファイルと同様に扱い、タギングのみで管理している。

管理ソフトは Apple の iTunes (図-3) である。iTunes は音楽ファイルの管理ソフトだが、どういうわけか PDF ファイルも管理できる。理由は定かではないが、iPod を接続しても転送されるわけでもないの、音楽ファイルに混ぜて PDF 化した紙文書の管理にも活用している。

PDF ファイルを iTunes にドラッグアンドドロップすると、ライブラリに追加されるので、必要に応じてアーティスト名 (作者名) やジャンルに情報を入力する。アルバム名に作者の所属する組織を入力するなどのアレンジを加えると使いやすくなるだろう。iTunes ではコメント欄も検索対象なので、ここを任意のタギングに利用して、複数のキーワードを改行などで区切って入力しておく。あとはタグや作者で絞り込んで探し出せるようになるのは、Picasa の場合と同様である。



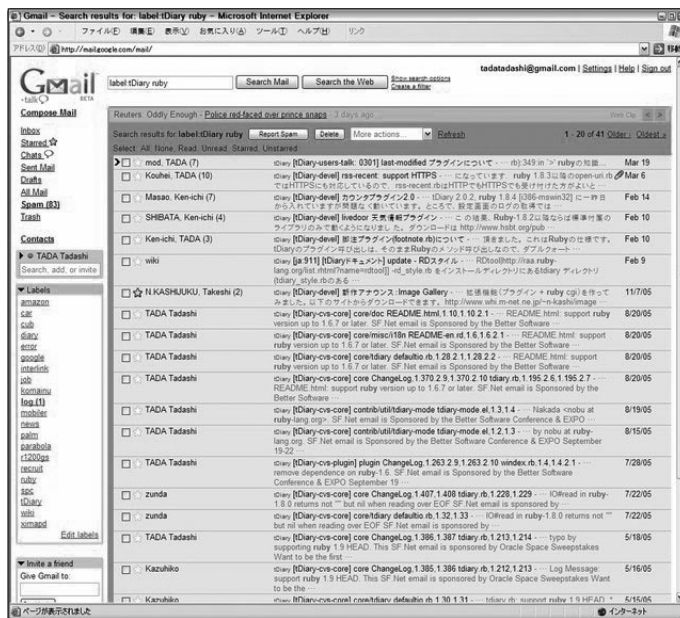


図-4 Gmail

## 電子メール

個人が管理するデータの中で、電子メールは参照頻度が最も高いものの1つだろう。電子メールはテキストデータなので、全文検索ソリューションで解決できることが多い。しかしながら、先に述べたようにランク付けのない全文検索は目的のデータを見つけ出すのに苦労するので、タギングで絞り込み可能にしておくのは有効だ。

筆者は Google の Gmail (図-4) を利用している。Google のサービスなので当然ながら全文検索機能が充実しているが、Picasa と同様にラベルと呼ばれる機能でタギングが可能である。

テキストデータにタギングを行う場合の利点は、そのテキストには含まれていない情報を付加できる点にある。過去の他人など、過去の自分同様に信用がならない。他人から届いたメール本文には含まれていない言葉をタグ

という形で付加しておけば、あとで検索するときに便利である。

Picasa, iTunes, Gmail と、管理アプリが複数あるのはあまり好ましい状況ではない。今後は Gmail の大容量を活かして、すべてのファイルを Gmail に送りつけることで整理ができないか、試してみるつもりだ。

### 参考 URL

- del.icio.us  
<http://del.icio.us/>
- はてなブックマーク  
<http://b.hatene.ne.jp/>
- MM/Memo  
<http://1470.net/mm/>
- Picasa  
<http://picasa.google.co.jp/>
- iTunes  
<http://www.apple.com/jp/itunes/>

(平成 18 年 4 月 3 日受付)

