

筑波大学大学院修士課程

理工学研究科修士論文

メタデータに基づく検索/閲覧インタフェース

佐藤 省吾

平成18年 3月

筑波大学大学院修士課程

理工学研究科修士論文

メタデータに基づく検索/閲覧インタフェース

佐藤 省吾

主任指導教員 システム情報工学研究科 田中 二郎

目次

論文要旨	1
第1章 はじめに	2
1.1 背景	2
1.2 分類と検索	2
1.3 研究の目的と本論文の構成	3
第2章 メタデータの利用	4
2.1 検索プロセス	4
2.2 検索結果の拡張	4
2.3 フィルタの適用による検索	5
2.3.1 データ集合	5
2.3.2 操作	5
2.3.3 操作の再利用	6
2.4 ビューの使い分けによる閲覧	6
第3章 検索/閲覧システム dripdrop	8
3.1 システムの概要	8
3.2 検索条件の編成	9
3.2.1 フィルタの追加	9
3.2.2 パラメータの調整	13
3.2.3 有効/無効の切り替え	13
3.2.4 フィルタの削除	14
3.2.5 フィルタの保存と読み込み	14
3.3 表形式ビュー	14
3.4 時系列ビュー	14
3.4.1 時刻の移動	15
3.4.2 時間単位の切り替え	15
第4章 dripdrop の実装	18
4.1 システム構成	18
4.1.1 データ管理部	18
4.1.2 フィルタリング部	19

4.1.3	ビュー部	19
4.2	検索対象のデータ	19
4.3	システム利用例	20
第 5 章	関連研究/関連システム	30
5.1	情報の管理と検索	30
5.2	多次元属性情報の検索/閲覧	31
5.3	時系列に基づく視覚化	31
第 6 章	議論	33
6.1	提案手法について	33
6.1.1	検索パラメータの指定	33
6.1.2	緩めるフィルタ	33
6.2	試作システムについて	34
6.2.1	データの取得	34
6.2.2	インタフェース	34
第 7 章	おわりに	35
	謝辞	35
	参考文献	36

目 次

1.1	分類主体の管理手法	3
1.2	検索主体の管理手法	3
2.1	検索結果の拡張	4
3.1	システム概観	8
3.2	フィルタ	9
3.3	検索条件の編成	9
3.4	絞り込むフィルタ	10
3.5	緩めるフィルタ	11
3.6	件数を制限するフィルタ	11
3.7	ドラッグ&ドロップによる直接選択	12
3.8	パラメータの調整	13
3.9	有効/無効の切り替え	13
3.10	フィルタの保存と読み込み	14
3.11	時系列ビュー	15
3.12	時刻の移動	16
3.13	時間単位の切り替え	17
4.1	システム構成	18
4.2	初期状態	20
4.3	フィルタの読み込み	21
4.4	フィルタの読み込み完了	22
4.5	キーワードの指定	22
4.6	パラメータの調整	23
4.7	ニュースサイトの履歴を選択	23
4.8	Web ページを開く	24
4.9	時系列ビューでの閲覧	24
4.10	日付について並べ替え	25
4.11	直接選択による絞り込み	25
4.12	直接選択による絞り込み完了	26
4.13	時間帯指定による抽出 (a)	26

4.14 時間帯指定による抽出 (b)	27
4.15 時間帯指定による抽出完了	27
4.16 データの種類を制限 (a)	28
4.17 データの種類を制限 (b)	28
4.18 セミナーの資料を選択	29
4.19 ドキュメントを開く	29

表 目 次

2.1	説明に用いる記号	5
4.1	検索対象のデータ	20

論文要旨

本研究では、メタデータを利用して個人の情報を検索/閲覧するシステムを試作した。

このシステムでは、検索条件を構成する個々の操作をフィルタとして捉え、フィルタを繰り返し適用することにより段階的にデータを取り出していく手法を採っている。絞り込んだ結果に対して条件を緩めることにより、異なる観点から結果を拡張できることを特徴としている。また、時刻情報に注目してデータを閲覧する手法を提案し、ビューの試作を行った。

以上の提案手法を踏まえてシステム“dripdrop”を作成し、メタデータに基づく検索/閲覧を実現した。

第1章 はじめに

1.1 背景

計算機が一般の家庭に普及し、多くの人々に使われるようになった。処理性能の向上やストレージ容量の増加に伴い、日常生活で利用するさまざまな情報を計算機で管理する機会が増えている。このような情報は種類や量が多く、また頻繁に変化するという特徴がある。近い将来、個人が計算機によって扱う情報はますます多くなると予想され、それらを管理する手法が重要性を増している。

実世界においてもものを整理しようとする場合、似たものをグループとしてまとめ、これを繰り返すことによって分類を進める方法が多くの場面で行われていると考えられる。この管理手法は階層構造に基づくものであり、ファイルシステムなどにも見られるように計算機における情報管理にも広く使われている。

この手法は、日常生活の行動に近い分りやすい、比較的簡単なデータ構造であるためプログラムによる実現が容易であるといった利点がある。しかし、このような整理方法は分類を前提としているため、対象の種類や量が多くなるとうまく機能しない場合がある。階層構造によって整理された情報に対して検索を行う場合、ユーザは情報のある程度適切に分類しておかなければならないが、現実には存在する多種多様のデータをすべて適切に分類するのは困難である。また、分類された状態を常に保っておく必要があり、内容の変化に応じた分類項目のメンテナンスが欠かせない。こういった作業はユーザに対する負担が大きいため、情報を保存する際の分類よりも情報を利用する際の検索に重きを置いた情報管理手法が現れてきている [1, 2, 3]。

1.2 分類と検索

分類主体の管理手法および検索主体の管理手法の特徴を、それぞれ図 1.1、図 1.2 に示す。

分類主体の手法では、階層構造に基づいて整理された分類項目が先に存在する。ここに新しくデータを追加する場合、ユーザはまずそのデータが対応する分類項目を判断し、その位置に格納する。データを取り出す場合についても、同じく最初に対応する分類項目を探し出し、そこからデータを取り出すという流れになる。

データと分類項目の対応付けが明確な状況下では優れた管理手法になりうるが、1つのデータが同時に複数の分類項目に所属することは原則として不可能であるため、格納時に不適切な分類項目に対応付けてしまうと検索が困難になるという欠点がある。

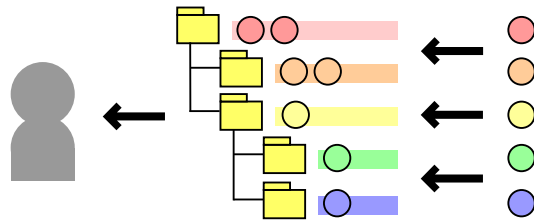


図 1.1 分類主体の管理手法

一方，検索主体の手法では個々のデータが先に存在する．そして，種類や性質によって分けることなくすべてのデータを一元的に扱う．このためデータを追加する場合であれば分類項目との対応を考える必要がなくなり，またデータを取り出す場合であれば特定の分類方法に縛られることがなくなる．検索条件の指定を変化させることによって，さまざまな観点から必要とするデータを取り出すことができる．

分類主体の手法に比べると実現する際のコストは高くなるが，個人が日常利用するような雑多な情報はある 1 つの観点から適切に分類することが困難であるため，検索主体の管理手法の方が適していると考ええる．

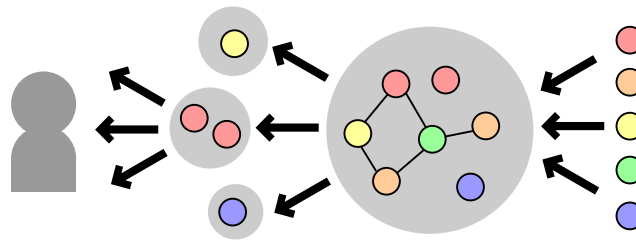


図 1.2 検索主体の管理手法

将来，個人が所有する多種・多量のデータに対して検索主体の管理手法を適用した場合，ユーザが希望する情報を取り出すための手法は重要な役割を持つことが予想される．

1.3 研究の目的と本論文の構成

個人の所有する情報が著しく増加しつつある背景を踏まえ，多種・多量の情報の中から希望するものを効率良く取り出すインタフェースを開発することを本研究の目標とする．

本論文の構成を以下に示す．

第 2 章では，メタデータに基づくフィルタリング手法を提案する．第 3 章では試作したシステムについて説明し，続く第 4 章でその実装について使用例を交えて述べる．第 5 章では関連研究および関連システムについて言及する．第 6 章で提案手法や試作システムに対して考察を加え，最後に第 7 章で結論を述べる．

第2章 メタデータの利用

2.1 検索プロセス

分類主体の管理手法では、情報を保存するための手段と取り出すための手段が同一であることが多かった。すなわち階層的に情報を保存し、それと同じ階層を通して情報を取得していた。しかし、検索主体の管理手法では両者は明確に区別されている。すべての情報は一箇所に保存されており、その中から必要なものだけを新たに取り出す方法を考える必要がある。

Teevan らの実験によると、メールやファイル、Web などの構造化されていない情報に対して検索を行う場合、キーワードを使って直接対象を見つけ出すよりも、小さなステップの繰り返しによって対象に近づいていく方法が好まれるという [4]。

本研究においても、単純な条件を積み重ねながらインタラクティブに検索を進めることを基本方針とする。このとき個々の情報そのものについての情報、すなわちメタデータを検索条件に利用する。

2.2 検索結果の拡張

検索によってある程度結果を絞り込んだ後、その結果と関連のある別のデータを取得したくなる場合がある。また、探しているデータを直接絞り込むのは難しいものの、それに近いデータを取得することなら可能である場合がある。絞り込みを行うだけの検索手法ではこのような状況にこれ以上対処できないため、絞り込んだ結果をもとにデータ集合を拡張する手法を提案する。

図 2.1 はこの手法を図示したものである。大きい円がデータの全体集合を、小さい円が個々のデータを表し、グレーの領域に含まれているデータが絞り込みの結果であるとする。まず、最初の状態から絞り込みを行い、一部のデータを取得する (a)。さらに絞り込みを進め (b)、その結果を別の観点から広げている (c)。この (c) の操作が検索結果の拡張に対応する。

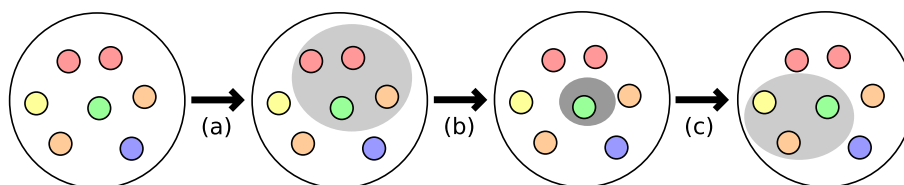


図 2.1 検索結果の拡張

2.3 フィルタの適用による検索

希望する結果を得るための検索条件が最終的には複雑なものになる場合であっても，検索を進める途中の個々の操作は，条件の追加やデータの並べ替えといった単純なものである．そこでこれらの操作をデータ集合に対するフィルタと捉え，フィルタを段階的に組み合わせることにより複雑かつ多様な検索を可能にする．また，作成したフィルタを保存しておくことで検索条件を再利用することができる．

フィルタによる検索手法は，データ集合とそれに対する操作から構成される．

2.3.1 データ集合

本研究で扱うデータ集合は，その要素として複数種のデータを同時に持つことができ，かつ要素の処理順序を保持しているものとする．集合に含まれる個々のデータには名前や日時などのメタデータが付けられており，各データはメタデータを共有することによって緩やかに関連を持っていると考えられる．

2.3.2 操作

データ集合に対して行う操作である．操作はデータ集合を受け取り，データ集合を返す関数であるといえる．また各操作の処理内容は，与えられたデータ集合から順番にデータを取り出し，条件を満たすデータのみを集めた集合を返すというものになる．

本研究で提案する操作を以下に示す．なお，説明の中で表 2.1 の記号を用いる．

表 2.1 説明に用いる記号

記号	説明
S_0	直前のデータ集合
U	全体のデータ集合
x	データ
p_i	パラメータ
$predicate$	データを受け取り，結果に含めるか否かを判定する関数
$predicate_generator$	パラメータを受け取り，判定用の関数を返す関数

絞り込む 直前のデータ集合のサブセットを取得する．絞り込む操作 f_n は次のように定式化できる．

$$\begin{aligned} predicate &= predicate_generator(p_0, \dots, p_n) \\ S_0 \supseteq f_n(S_0) &= \{x \mid x \in S_0, predicate(x)\} \end{aligned}$$

緩める 直前のデータ集合を何らかの観点に基づいて拡張し、スーパーセットを取得する。すなわち全体集合のサブセットを取得し、直前のデータ集合との和集合を返す操作である。このときの条件は直前のデータ集合を用いて定義される場合があるため、緩める操作 f_w は次のように定式化できる。

$$\begin{aligned} predicate &= predicate_generator(S_0, p_0, \dots, p_n) \\ S_0 \subseteq f_w(S_0) &= S_0 \cup \{x \mid x \in U, predicate(x)\} \\ &= S_0 \cup f_n(U) \end{aligned}$$

並べ替える データ集合をある観点のもとに並び替え、後続の操作によって処理される順番を変更する。操作によってはこの順番が結果のデータ集合に影響する場合がある。

たどる データ集合に含まれる各データに対して、関連付けられた他のデータを取り出す。

件数を制限する 取り出すデータ数を制限する。

2.3.3 操作の再利用

よく使う操作を保存しておき、必要に応じて修正を加えて利用すると便利であると考える。例えば以下の一連の操作は、Google で “Java” を検索した結果からたどったページを抽出するものである。

- Web の閲覧履歴のうち
- タイトルに “Google 検索” を含み
- ホスト名に “google” を含み
- さらにタイトルに “Java” を含むページから
- 5 ページ以内に見たもので
- 1 分以上表示していたものを
- 新しい順に並べる

これに対応するフィルタを作成し保存しておけば、保存したフィルタを読み込んで “Java” の箇所を変更することによりで他の検索結果にも適用できる。また、さらに条件を追加して検索を進めることも可能となる。

2.4 ビューの使い分けによる閲覧

検索結果を閲覧する際、注目するメタデータに応じて複数のビューを使い分ける。例えば、時間情報にはカレンダー形式のビューを、位置情報には地図形式のビューを、詳細を見たい

場合は表形式のビューを用いるといった使い分けが考えられる。

本研究ではメタデータの1つである時刻に注目したビューを提案する。時刻は多くのデータに付随する一般的な属性であり、時刻に注目することによってさまざまなデータを一次元上に並べ、眺めることができる。

時系列の特性として以下の2つが考えられる。

- 連続性
過去から未来に向かって一次元に変化する。
- 周期性
月，日，時など一定の周期に基づいて捉える。

これらの特性を反映したビューとするため、円筒の表面に対して時間軸を螺旋状に巻きつけたモデルを考える。円筒を横方向から見た場合の矩形が連続的な視点に、また上下方向から見た場合の円が周期的な視点に対応する。

第3章 検索/閲覧システム dripdrop

3.1 システムの概要

第2章で提案した手法に基づき、情報検索/閲覧システム“dripdrop”を試作した。

dripdrop の概観を図 3.1 に示す。画面は左右に分割された2部構成となっており、左側が検索条件編成部、右側が検索結果表示部である。検索条件に変更を加えると即座に検索結果の表示に反映される。検索結果は、表形式のビューまたは時系列に基づくビューによって閲覧する。図では表形式のビューを使用している。

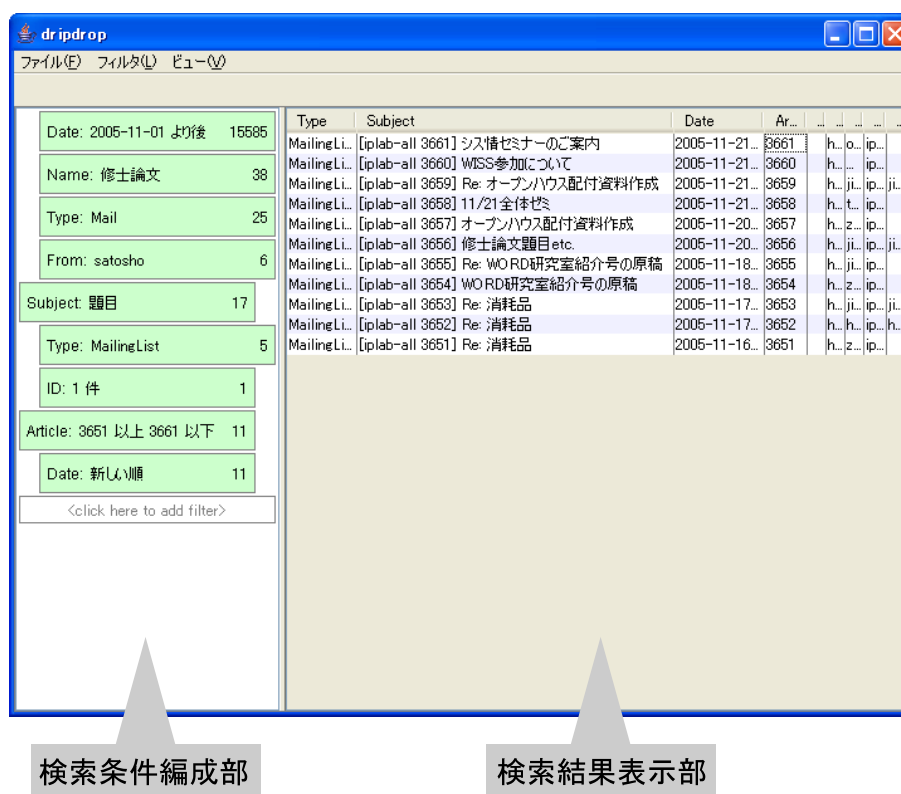


図 3.1 システム概観

3.2 検索条件の編成

検索条件編成部において、個々のフィルタは1つの矩形として表現される（図 3.2）。矩形内部の表示は左からフィルタの種類、フィルタリング動作のパラメータ、フィルタ適用後のデータ数を意味する。編成部全体に対する水平方向の表示位置はフィルタの種類によって異なり、絞り込む、緩める等の操作を一目で判断できるようにしている。

検索条件の編成を進める流れを図 3.3 に示す。(a) では2つのフィルタが適用されている。ここから (b)(c) の手順を経ることによりフィルタを1つ追加した状態が (d) である。このようにしてフィルタの追加を繰り返すと (e) のような状態に至る。データ全体に対して上から順にフィルタが適用され、最終的な結果が検索結果として表示される。

以下、検索条件の編成に関する個々の機能について説明する。

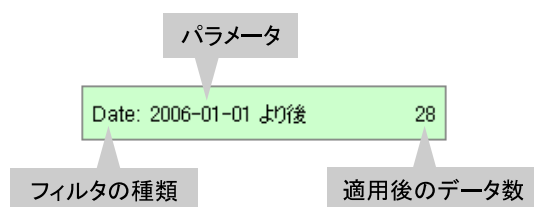


図 3.2 フィルタ

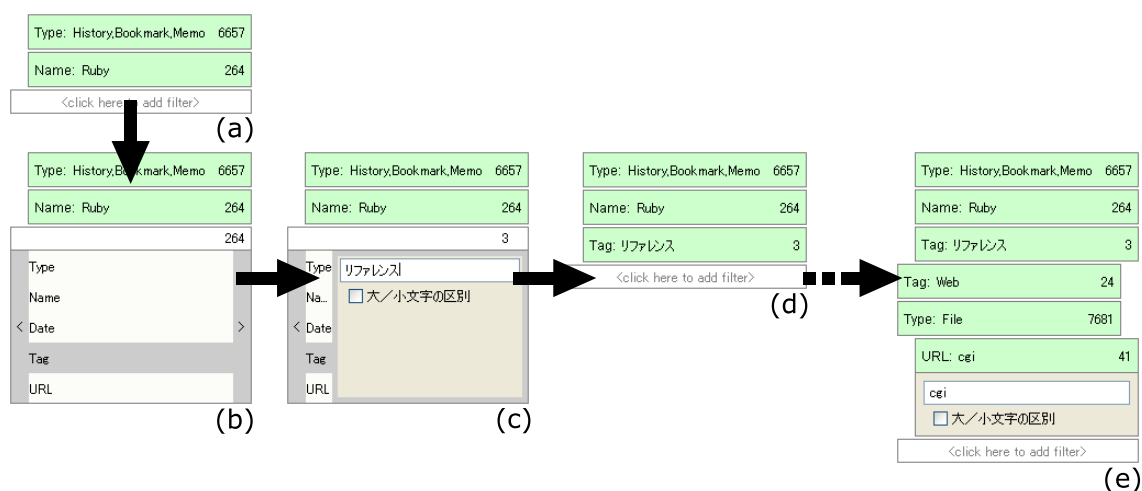


図 3.3 検索条件の編成

3.2.1 フィルタの追加

フィルタを追加する方法はいくつか存在し、フィルタの種類によって異なる。

プロパティに対するフィルタ

プロパティに対して何らかの操作を行うフィルタを追加するには、編成部下端の<click here to add filter>と表示されている箇所をクリックする。するとその時点で条件の作成に利用可能なプロパティが計算され、ドロップダウンリストに一覧表示される。

ユーザはリスト中の各プロパティを利用して以下の操作を行うことができる。

- 絞り込む
選択したプロパティについて条件を指定し、その条件を満たすデータを直前のデータ集合から抽出する。
- 緩める
選択したプロパティについて条件を指定し、その条件を満たすデータを全体のデータ集合から抽出する。その結果を直前のデータ集合に加える。
- 並び替える
選択したプロパティをキーとしてデータを並べ替え、データ集合におけるデータの処理順序を変更する。

リスト上では、項目の上にマウスポインタを移動するとその項目が選択される。「絞り込む」操作を行う場合、利用するプロパティを選択してから右側にマウスポインタを移動すると、パラメータ設定用のパネルが開いて条件を指定することができる。条件指定後、パネルの下側にマウスポインタを出すことにより操作が確定し、フィルタが追加される（図 3.4）。このとき下以外の方向に出すと操作の取り消しとなる。

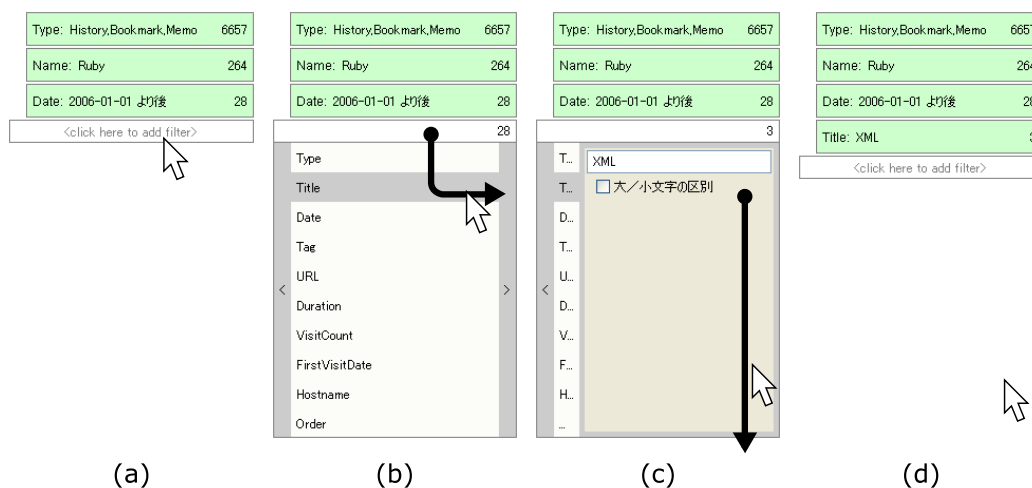


図 3.4 絞り込むフィルタ

「緩める」操作は、プロパティを選択してから左側にマウスポインタを移動する点を除き、絞り込む操作と同様である（図 3.5）。

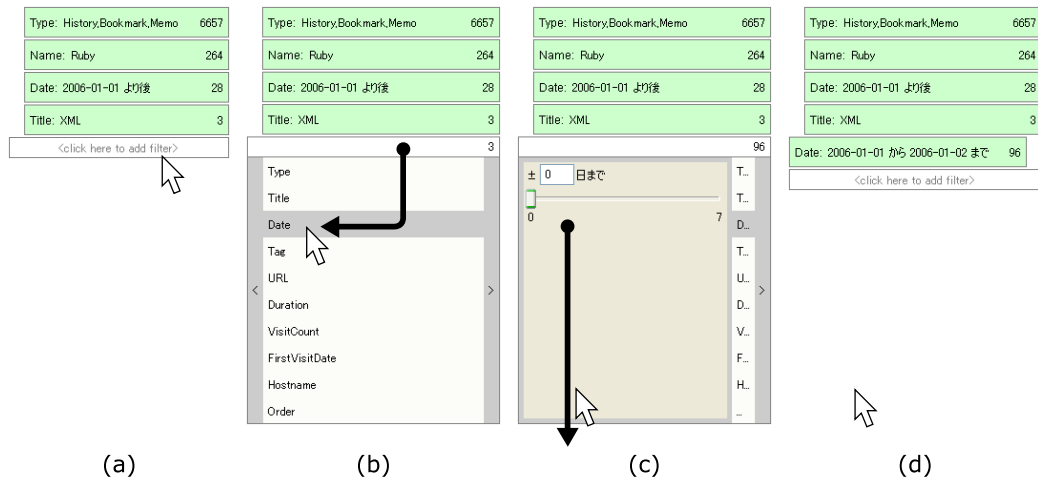


図 3.5 緩めるフィルタ

また「並べ替える」操作を行う場合は、プロパティ選択後に項目上で左クリックすると対応するフィルタが追加される。

データ数に基づくフィルタ

検索結果の件数を制限するフィルタはアプリケーションのメニューから追加する。「先頭から件数制限」または「末尾から件数制限」メニューを実行すると、対応するフィルタが追加される。図 3.6 にフィルタを追加した後の状態を示す。



図 3.6 件数を制限するフィルタ

選択に基づくフィルタ

表形式のビューにおいて抽出したい項目を選択し、そのまま編成部にドラッグ&ドロップすると、選択したデータのみを直接抽出するフィルタを追加できる。図 3.7では、ドラッグ&ドロップによって ID という種類のフィルタが追加されていることを確認できる。

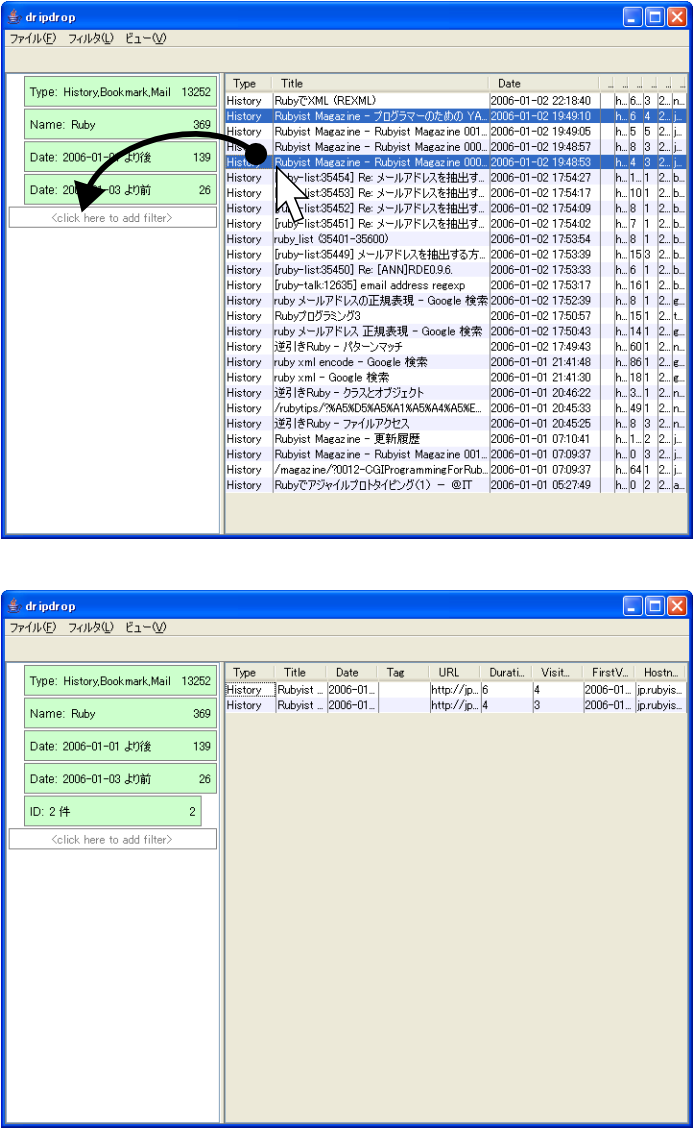


図 3.7 ドラッグ&ドロップによる直接選択

3.2.2 パラメータの調整

フィルタを左クリックすると設定用のパネルが開き、パラメータの調整ができる（図 3.8）。パネルの内容はフィルタごとに異なり、テキスト入力フィールドやチェックボックス、スライダなど設定項目に応じたインタフェースが用意されている。

なお、「絞り込む」「緩める」フィルタを追加する際に表示されるパネルはここで説明したものと同一である。

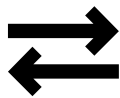
Type: History,Book mark,Memo	6657
Name: Ruby	264
Date: 2006-01-01 より後	28
Title: XML	3
XML	
☐ 大／小文字の区別	
Date: 2006-01-01 から 2006-01-02 まで	96
<click here to add filter>	

図 3.8 パラメータの調整

3.2.3 有効/無効の切り替え

フィルタを右クリックするとそのフィルタが一時的に無効になり、もう一度右クリックすると元に戻る（図 3.9）。無効になったフィルタは背景色が暗くなり、すべてのデータを無条件で通すようになる。こうして有効/無効を切り替えることにより、そのフィルタの働きを確認することができる。

Type: History,Book mark,Memo	6657
Name: Ruby	264
Date: 2006-01-01 より後	28
Title: XML	3
Date: 2006-01-01 から 2006-01-02 まで	96
<click here to add filter>	



Type: History,Book mark,Memo	6657
Name: Ruby	6657
Date: 2006-01-01 より後	634
Title: XML	17
Date: 2006-01-01 から 2006-01-03 まで	17
<click here to add filter>	

図 3.9 有効/無効の切り替え

3.2.4 フィルタの削除

不要なフィルタの上でマウスの中ボタンをクリックすると、そのフィルタが削除される。

3.2.5 フィルタの保存と読み込み

メニューから、現在編成しているフィルタのリストをファイルに保存したり、ファイルに保存されているフィルタのリストを読み込んだりすることができる（図 3.10）。これにより過去に作成した検索条件を再利用することが可能になる。ファイルから読み込んだフィルタは、通常どおり追加/削除やパラメータ調整の対象となる。

ファイル(F)	フィルタ(L)	ビュー(V)
	先頭から件数制限(H)	
	末尾から件数制限(T)	
Type:	フィルタをロード(L)...	6657
Name	フィルタを保存(S)...	264
	すべて削除(A)	
Date: 2006-01-01 より後		28

図 3.10 フィルタの保存と読み込み

3.3 表形式ビュー

検索結果表示部で利用可能なビューの1つは表形式のものである（図 3.1）。このビューは一般的なスプレッドシート型インタフェースと同様、各行が1つのデータに、各列が1つのプロパティに対応する。すべてのデータに共通するプロパティのみが列として表示される。

また、表のヘッダをクリックするとそのプロパティをキーとしてソートを行う機能や、行をダブルクリックすると対応するアプリケーションでそのデータを開く機能も持たせている。

3.4 時系列ビュー

検索結果表示部で利用可能なもう1つのビューである。時間軸を表す縦長の矩形と、時間単位を表すリングから構成される（図 3.11）。ユーザはリングの上でマウスをドラッグすることにより閲覧する時間帯を設定する。各データは、それぞれの持つ時刻に応じて時間軸上の対応する位置に表示される。なお時間軸は下向きを正とする。

リングの左下に表示されている文字列はリングが表現する時間単位である。リングの上方には現在注目している時刻が表示されており、この時刻は線で連結された矩形の中央部に対応する。また、矩形の中央付近で色が濃くなっている領域は、リングが表している時間単位と一致する。図 3.11 を例に挙げると“2005 年 12 月 3 日 12 時 41 分 55 秒を中心とする 1ヶ月間”となる。

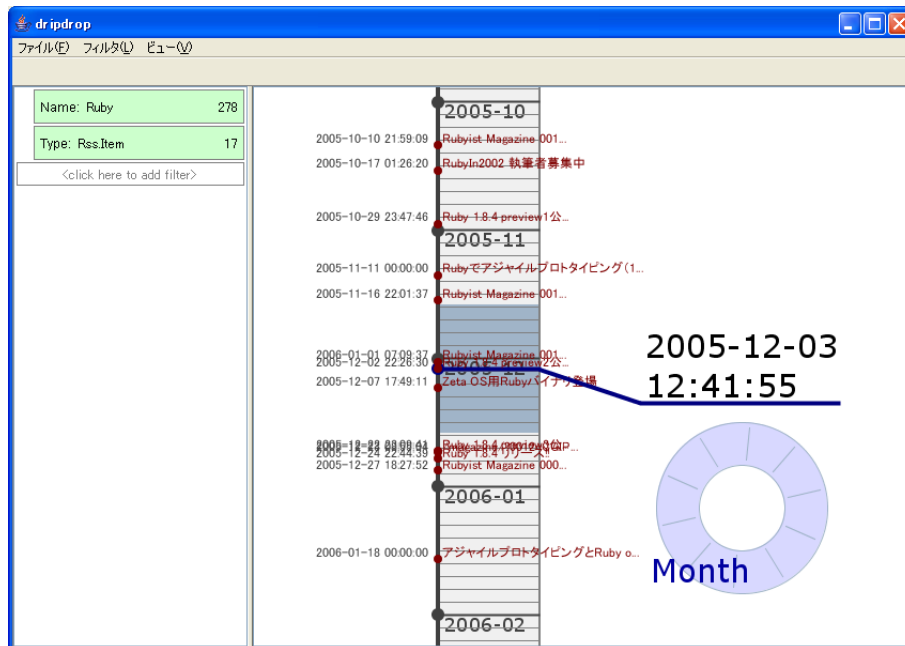


図 3.11 時系列ビュー

このビューでは、矩形・リング以外の場所をドラッグすることにより視点の変更が可能である。マウスの左ボタンがパニング、右ボタンがズーミングに対応している。

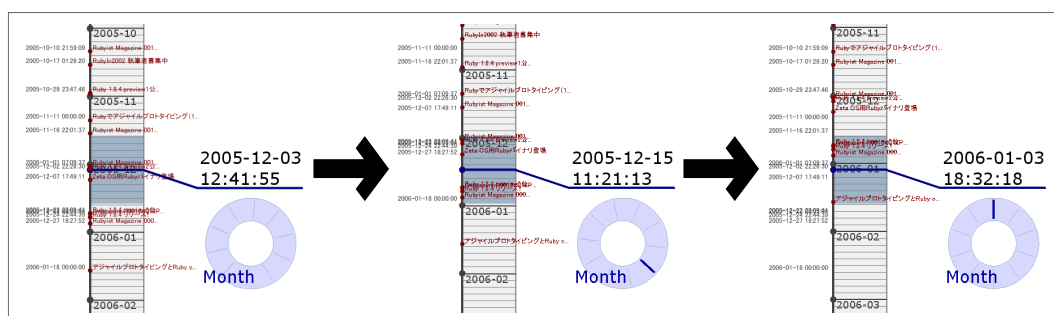
以下、リング上でのマウス操作で時間帯を変更する方法について説明する。

3.4.1 時刻の移動

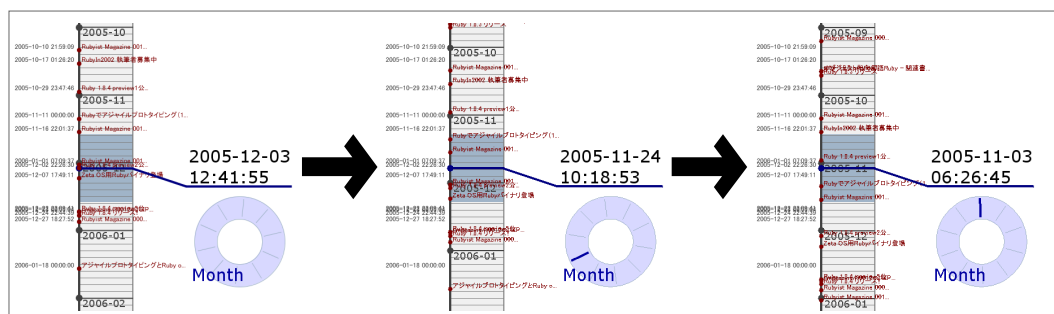
時間単位はそのまま時刻だけを移動する場合は円周方向に操作を行う。リングの上を時計回りにドラッグすると時刻が進み、反時計回りにドラッグすると戻る。1周が1単位に対応しており、図 3.12 (a) (b) ではそれぞれ前後約1ヶ月の移動を行っている。なお、時刻の移動は相対的な角度に基づいて行われるため、円周上のどの位置からドラッグを開始しても構わない。

3.4.2 時間単位の切り替え

時間単位を切り替える場合は半径方向に操作を行う。リングの上でドラッグを開始し、リングの外側にマウスポインタを移動すると単位が1段階大きくなる。同様に、内側に移動すると単位が1段階小さくなる。図 3.13 (a) (b) では“1ヶ月”を基準としてそれぞれ大小に単位を切り替えている。リングにおける単位の切り替えに伴い、矩形の時間単位も変化する。時刻の移動と同様ドラッグの開始位置はどこでも構わない。

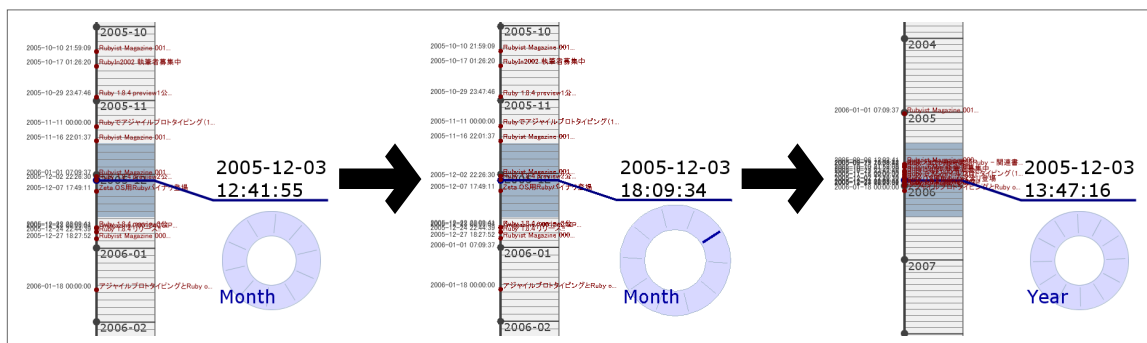


(a) 時刻を進める

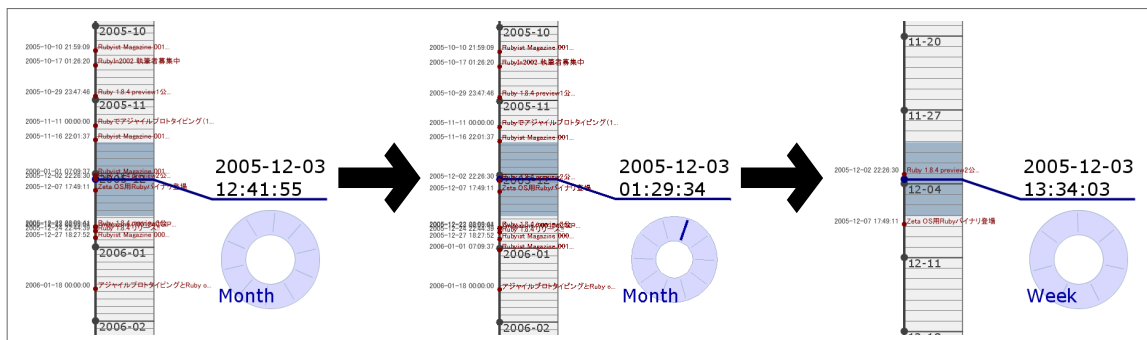


(b) 時刻を戻す

図 3.12 時刻の移動



(a) 時間単位を大きくする



(b) 時間単位を小さくする

図 3.13 時間単位の切り替え

第4章 dripdrop の実装

本章では、dripdrop の実装における詳細について説明する。

システムの開発には Java™ 2 Platform Standard Edition Development Kit 5.0 を用い、デスクトップ上で動作する Java アプリケーションとして実装した。

4.1 システム構成

システムの構成を図 4.1 に示す。試作システムは、データ管理部、フィルタリング部、ビュー部の3部から構成される。

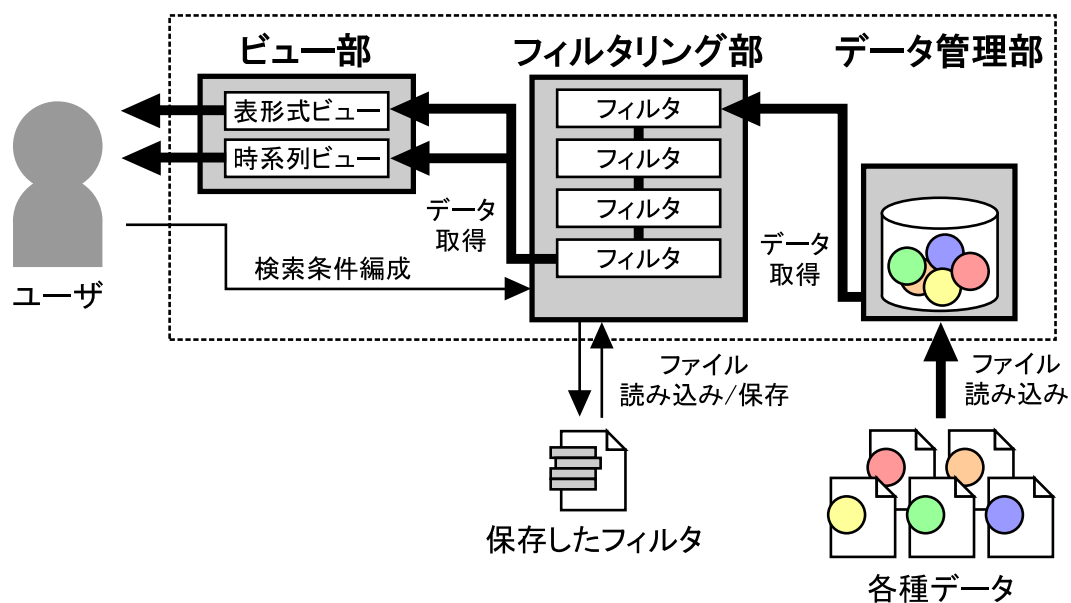


図 4.1 システム構成

4.1.1 データ管理部

データ管理部では複数種のデータを一元的に管理している。

個々の情報が持つメタデータは、W3C(World Wide Web Consortium)[5]によって標準化された RDF(Resource Description Framework)[6] と呼ばれる枠組みに基づいて表現している。RDF は Semantic Web を構成する要素技術の 1 つであり、特に Web 上のリソースに対してメタデータを記述する用途を意図して設計されている。RDF のモデルは有向ラベル付きグラフであるため自由度の高い記述が可能となっており、RSS(RDF Site Summary)[7] や FOAF(Friend of a Friend)[8] を始めとしたさまざまな応用がなされている。本システムでは、メタデータを記述する手段として一般的であること、また汎用性の高い形式であることから実装に RDF を用いた。

RDF モデルを XML で記述するための構文として RDF/XML が定義されている。データ管理部は、この構文に基づいてデータが記述された XML 文書を読み込み、RDF モデルとして内部に保持する。これらの機能を実現するため、Java で記述された Semantic Web のフレームワークである Jena 2.3[9] を利用した。

4.1.2 フィルタリング部

フィルタリング部では、データ管理部によって管理されているデータに対してフィルタリングを行い、条件を満たすデータを抽出する。また、ユーザによる条件の編成を受け付ける。

フィルタリング部の内部では、適用されているフィルタ群を一次元のリストとして保持している。あるフィルタに対して条件の変更が行われた場合、そのフィルタおよび下流にあるすべてのフィルタでデータの更新を行い、最終的にビューに通知する。

フィルタリングに使用できるプロパティはデータの種類に応じて決定される。データ集合に複数種のデータが含まれている場合は、すべての種類に共通するプロパティのみが対象となる。それぞれのプロパティに対して対応するフィルタが定義されており、またそれぞれのフィルタに対して対応するパラメータ設定パネルが定義されている。

4.1.3 ビュー部

ビュー部では、フィルタリング部によって抽出されたデータ集合をユーザに提示する。表形式ビューの実装には Java の開発キットに標準で付属している GUI コンポーネントを用いた。一方、時系列ビューの実装には Piccolo.Java 1.2[10] を利用した。Piccolo はズームインインタフェースの構築をサポートするツールキットである。

4.2 検索対象のデータ

試作システムで動作確認を行う際に検索対象としたデータを表 4.1 に示す。これらのデータは著者が日常的に使用しているアプリケーションから取り出したものである。それぞれに対応するスクリプト等を用意して RDF/XML 形式に変換し、ファイルに保存している。ただし、RSS 1.0 は RDF/XML で記述されているためそのままシステムに読み込むことができる。

表 4.1 検索対象のデータ

データの種類	件数	説明
履歴	10206	Web ページの閲覧履歴
ブックマーク	275	Web ページのブックマーク
メール	825	個人で受け取ったメール
メーリングリスト	1946	研究室のメーリングリスト（3 種類）
メモ	59	テキスト形式のメモ
RSS (channel)	14	RSS フィードの提供元
RSS (item)	1705	RSS フィードの項目
ファイル	22775	ホームディレクトリ以下にあるファイルの情報

4.3 システム利用例

ここでは、実際にシステムを利用して検索/閲覧を行う様子を説明する。検索の対象は表 4.1 に示した計 37,805 件のデータである。

図 4.2 はデータを読み込んだ直後の状態である。この時点ではまだフィルタが適用されていないため、左側の検索結果編成部は空になっている。この状態から検索および閲覧を開始する。

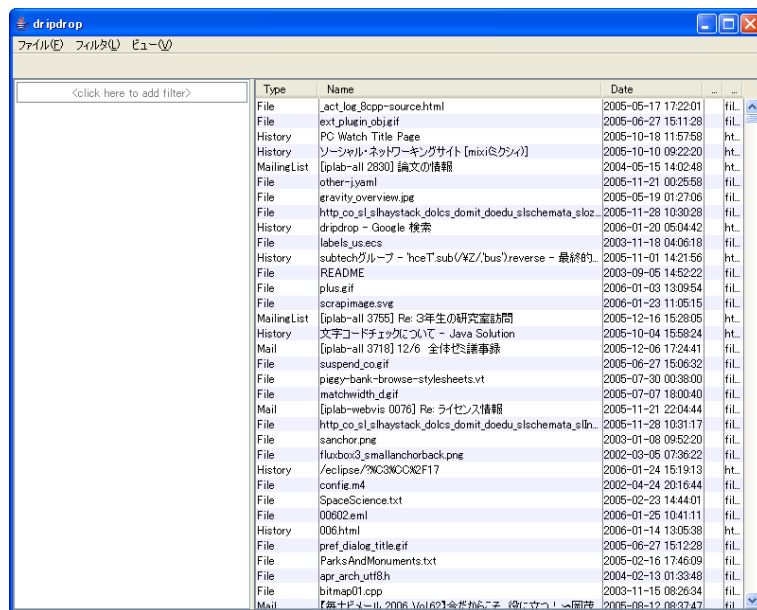


図 4.2 初期状態

例として次の状況を考える。

- ソーシャルブックマークサービスの“del.icio.us”について過去に Web で調べていた際、あるニュースサイトの記事を見かけたことを思い出した
- その記事には Google の検索結果からリンクをたどって到達したことを覚えている
- もう 1 度その記事を読みたいと思ったが URL が分からない
- ページのタイトルがよく思い出せない

この場合、探しているページは Web の閲覧履歴の中に存在する。しかしそのページを直接絞り込むための適切なキーワードを指定することが難しいため、Google で“del.icio.us”を検索した結果ページの履歴を取得し、そこからたどったページを調べていく方法を考えた。

もしこのような「Google の検索結果ページを取得する」という操作をしばしば行うのであれば、あらかじめフィルタを作成しておき、それを再利用することによって条件指定の手間を減らすことができる。図 4.3 では、過去に作成しておいたフィルタを読み込んでいる。

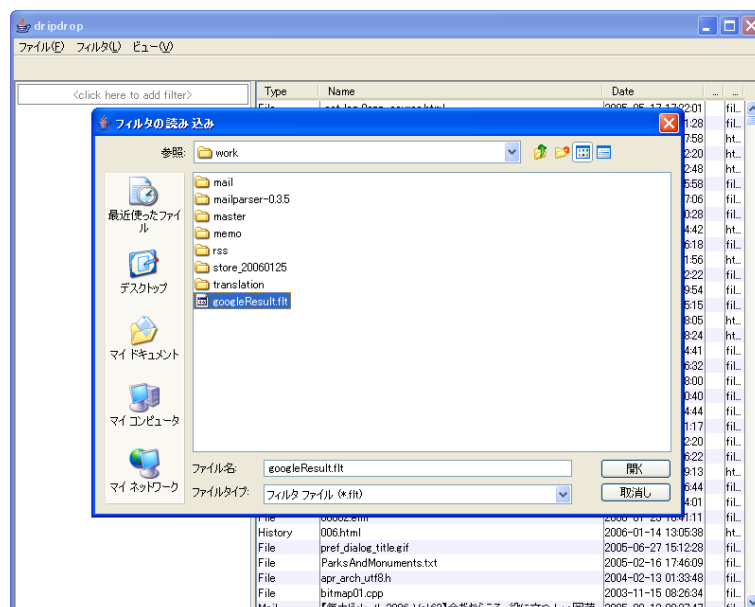


図 4.3 フィルタの読み込み

図 4.4 はフィルタの読み込みが完了した状態である。この一連のフィルタは、Web の閲覧履歴の中でタイトルに“Google 検索”を含み、かつホスト名に“google”を含むページ各々に対し、続く 3 ページ以内に見た Web ページを取得する。さらにその中から表示時間が 30 秒を超えていたもののみを抽出するという操作を行っている。以上の操作により Google 検索の結果ページを抽出することを想定している。

なお、上から 4 番目に組み込まれている Title フィルタはキーワードが空になっている。これは実際の使用時に値を設定するためである。

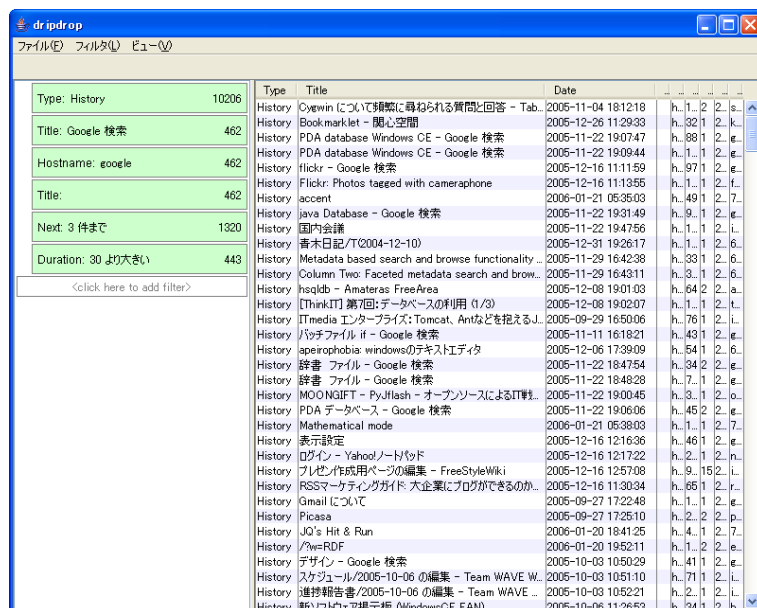


図 4.4 フィルタの読み込み完了

今回は“del.icio.us”を検索した結果を取得したいので、Title フィルタのキーワードに文字列“del.icio.us”を指定する（図 4.5）。

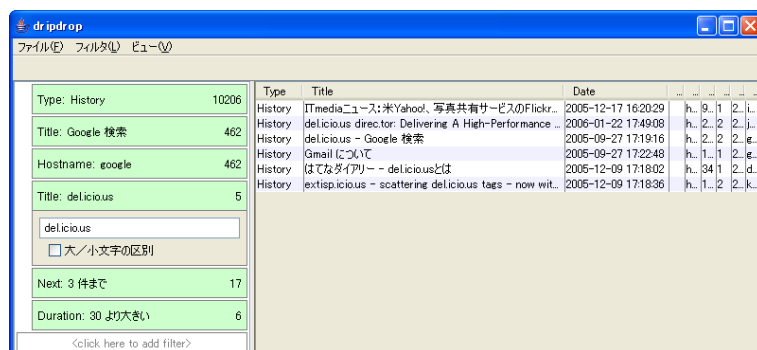


図 4.5 キーワードの指定

ここまでの操作により、10,206 件の履歴を 6 件まで絞り込むことができた。しかしこの結果の中に探しているページが含まれていなかったため、パラメータの調整を行い条件を緩くする。図 4.6 では、Next フィルタの件数を 3 件から 5 件に拡大し、表示時間の制限を一時的に無効にしている。

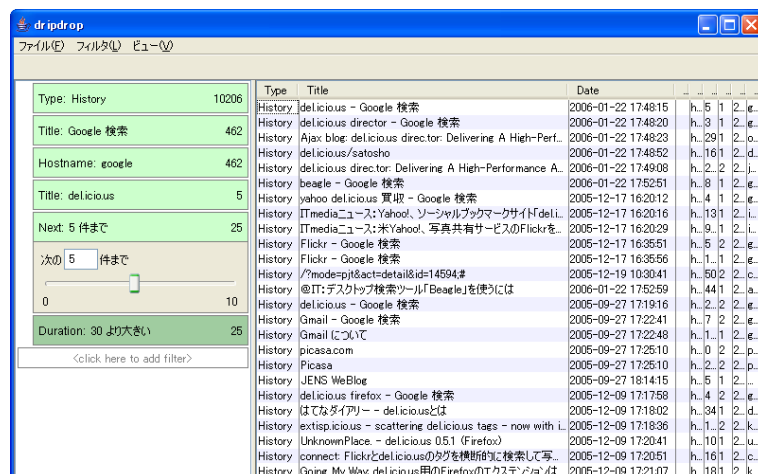


図 4.6 パラメータの調整

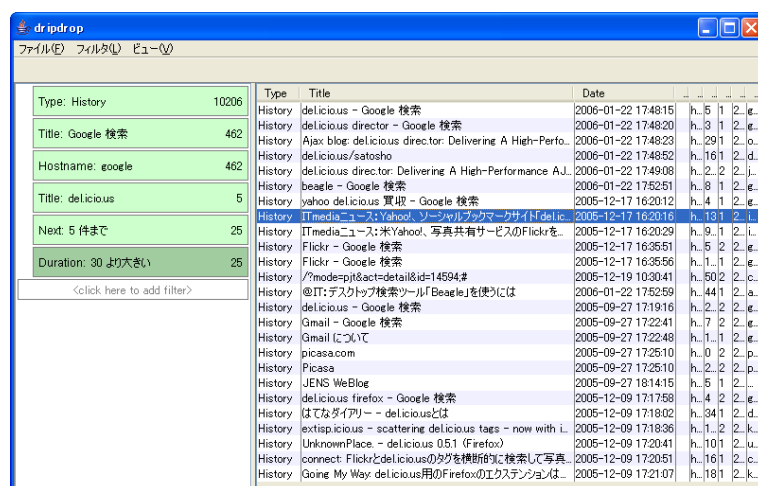


図 4.7 ニュースサイトの履歴を選択

パラメータを変更したことにより検索結果の件数が 25 件に増加した。結果のリストのタイトルを順に眺めていくと、探しているページと一致していそうなものを発見することができた (図 4.7)。ここで項目をダブルクリックして該当するページを開き、探しているページそのものであることを Web ブラウザ上で確認した (図 4.8)。



図 4.8 Web ページを開く

ここまでの操作により探していたページは見つかったが、図 4.7 の Date の列に注目すると、“del.icio.us”を検索している時期がその周辺以外にも存在することが分かった。これらが時系列上でどのような関係にあるか気になったため、時系列ビューへの切り替えを行う（図 4.9）。

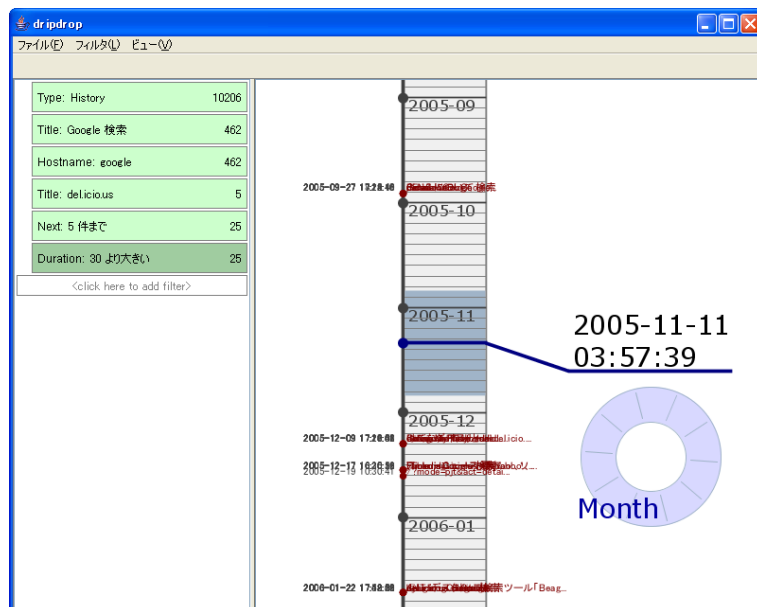


図 4.9 時系列ビューでの閲覧

時系列ビューで検索結果を見ると、9月下旬、12月上旬・中旬、1月下旬の4箇所で見られているようである。このうち、9月下旬にWeb検索をした頃は自分がどのような作業をしていたかに興味を持ち、その時期に関連する情報を見たくなった。まず現在の結果を日付について並べ替えた(図4.10)。

Type	Title	Date				
History	delicious - Google 検索	2005-09-27 17:19:16	☐	☐	☐	☐
History	Gmail - Google 検索	2005-09-27 17:22:41	☐	☐	☐	☐
History	Gmail について	2005-09-27 17:22:48	☐	☐	☐	☐
History	picasa.com	2005-09-27 17:25:10	☐	☐	☐	☐
History	Picasa	2005-09-27 17:25:10	☐	☐	☐	☐
History	JENS WeBlog	2005-09-27 18:14:15	☐	☐	☐	☐
History	delicious firefox - Google 検索	2005-12-09 17:17:58	☐	☐	☐	☐
History	はてなダイアリー - deliciousとは	2005-12-09 17:18:02	☐	☐	☐	☐
History	extispicious - scattering delicious tags - now wit...	2005-12-09 17:18:36	☐	☐	☐	☐
History	UnknownPlace - delicious 0.5.1 (Firefox)	2005-12-09 17:20:41	☐	☐	☐	☐
History	connect: Flickrとdeliciousのタグを横断的に検索して...	2005-12-09 17:20:51	☐	☐	☐	☐
History	Going My Way: delicious用のFirefoxのエクステンション...	2005-12-09 17:21:07	☐	☐	☐	☐
History	yahoo delicious 買取 - Google 検索	2005-12-17 16:20:12	☐	☐	☐	☐
History	ITmediaニュース: Yahoo!、ソーシャルブックマークサイト「de...	2005-12-17 16:20:16	☐	☐	☐	☐
History	ITmediaニュース: 米Yahoo!、写真共有サービスのFlickr...	2005-12-17 16:20:29	☐	☐	☐	☐
History	Flickr - Google 検索	2005-12-17 16:35:51	☐	☐	☐	☐
History	Flickr - Google 検索	2005-12-17 16:35:56	☐	☐	☐	☐
History	/?mode=pi&act=detail&id=14594#	2005-12-19 10:30:41	☐	☐	☐	☐
History	delicious - Google 検索	2006-01-22 17:48:15	☐	☐	☐	☐
History	delicious director - Google 検索	2006-01-22 17:48:20	☐	☐	☐	☐
History	Ajax blog: delicious director: Delivering A High-Per...	2006-01-22 17:48:23	☐	☐	☐	☐
History	delicious/satosho	2006-01-22 17:48:52	☐	☐	☐	☐
History	delicious director: Delivering A High-Performance...	2006-01-22 17:49:08	☐	☐	☐	☐
History	beagle - Google 検索	2006-01-22 17:52:51	☐	☐	☐	☐
History	@IT: デスクトップ検索ツール「Beagle」を使うには	2006-01-22 17:52:59	☐	☐	☐	☐

図 4.10 日付について並べ替え

9月下旬に相当する最初と最後の項目を選択し、条件編成部にドラッグ&ドロップすることにより直接選択を行う(図4.11)。

Type	Title	Date				
History	delicious - Google 検索	2005-09-27 17:19:16	☒	☐	☐	☐
History	Gmail - Google 検索	2005-09-27 17:22:41	☐	☐	☐	☐
History	Gmail について	2005-09-27 17:22:48	☐	☐	☐	☐
History	picasa.com	2005-09-27 17:25:10	☐	☐	☐	☐
History	Picasa	2005-09-27 17:25:10	☐	☐	☐	☐
History	JENS WeBlog	2005-09-27 18:14:15	☐	☐	☐	☐
History	delicious firefox - Google 検索	2005-12-09 17:17:58	☐	☐	☐	☐
History	はてなダイアリー - deliciousとは	2005-12-09 17:18:02	☐	☐	☐	☐
History	extispicious - scattering delicious tags - now wit...	2005-12-09 17:18:36	☐	☐	☐	☐
History	UnknownPlace - delicious 0.5.1 (Firefox)	2005-12-09 17:20:41	☐	☐	☐	☐
History	connect: Flickrとdeliciousのタグを横断的に検索して...	2005-12-09 17:20:51	☐	☐	☐	☐
History	Going My Way: delicious用のFirefoxのエクステンション...	2005-12-09 17:21:07	☐	☐	☐	☐
History	yahoo delicious 買取 - Google 検索	2005-12-17 16:20:12	☐	☐	☐	☐
History	ITmediaニュース: Yahoo!、ソーシャルブックマークサイト「de...	2005-12-17 16:20:16	☐	☐	☐	☐
History	ITmediaニュース: 米Yahoo!、写真共有サービスのFlickr...	2005-12-17 16:20:29	☐	☐	☐	☐
History	Flickr - Google 検索	2005-12-17 16:35:51	☐	☐	☐	☐
History	Flickr - Google 検索	2005-12-17 16:35:56	☐	☐	☐	☐
History	/?mode=pi&act=detail&id=14594#	2005-12-19 10:30:41	☐	☐	☐	☐
History	delicious - Google 検索	2006-01-22 17:48:15	☐	☐	☐	☐
History	delicious director - Google 検索	2006-01-22 17:48:20	☐	☐	☐	☐
History	Ajax blog: delicious director: Delivering A High-Per...	2006-01-22 17:48:23	☐	☐	☐	☐
History	delicious/satosho	2006-01-22 17:48:52	☐	☐	☐	☐
History	delicious director: Delivering A High-Performance...	2006-01-22 17:49:08	☐	☐	☐	☐
History	beagle - Google 検索	2006-01-22 17:52:51	☐	☐	☐	☐
History	@IT: デスクトップ検索ツール「Beagle」を使うには	2006-01-22 17:52:59	☒	☐	☐	☐

図 4.11 直接選択による絞り込み

図 4.12 では、選択した2つのデータのみが表示されている。

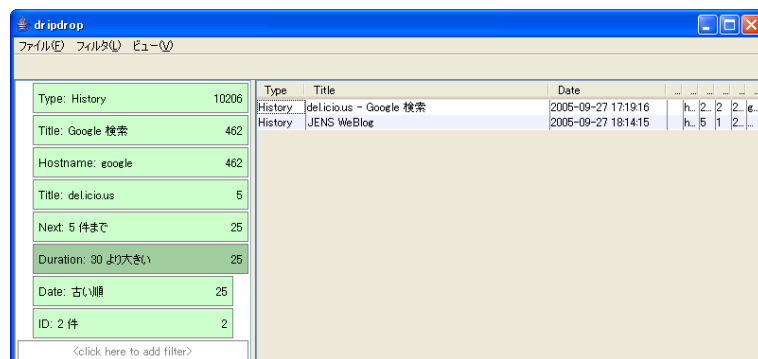


図 4.12 直接選択による絞り込み完了

表示時間に基づくフィルタは既に不要であると判断し、ここでリストから削除した。続いて、先程選択した2つのデータを両端とする時間帯に注目する（図 4.13, 4.14）。

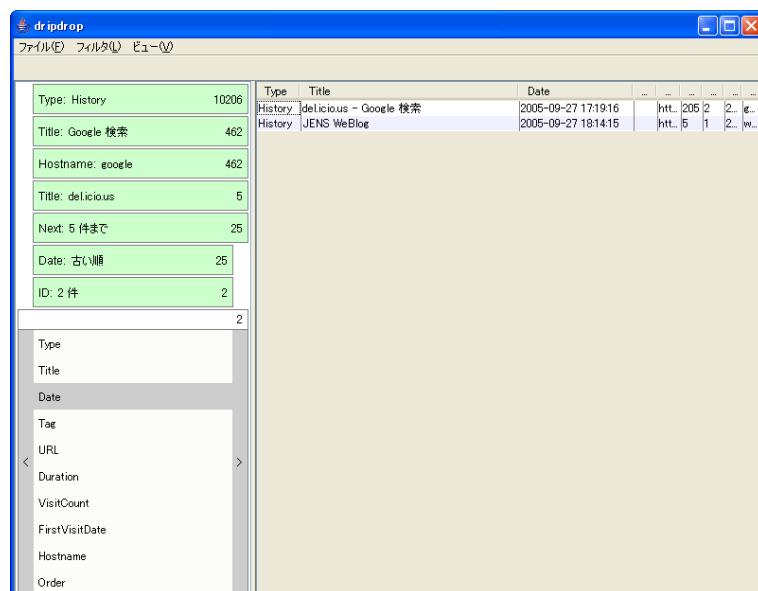


図 4.13 時間帯指定による抽出 (a)

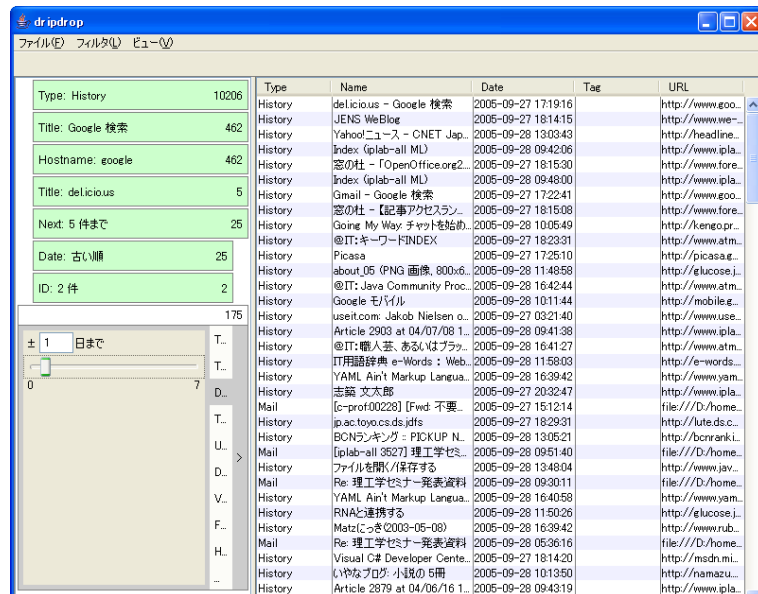


図 4.14 時間帯指定による抽出 (b)

さらに前後 1 日まで時間帯を拡張したところ、図 4.15 のような結果になった。

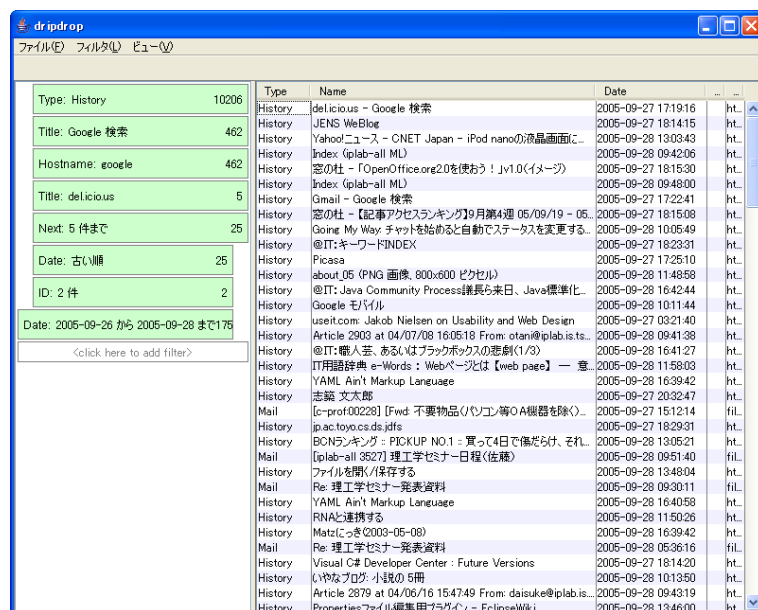


図 4.15 時間帯指定による抽出完了

続けてデータの種類をファイルに制限する（図 4.16, 4.17）.

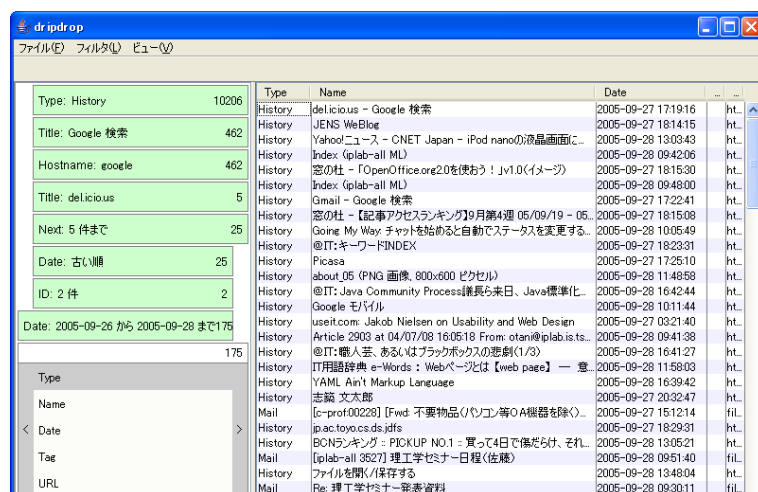


図 4.16 データの種類を制限 (a)

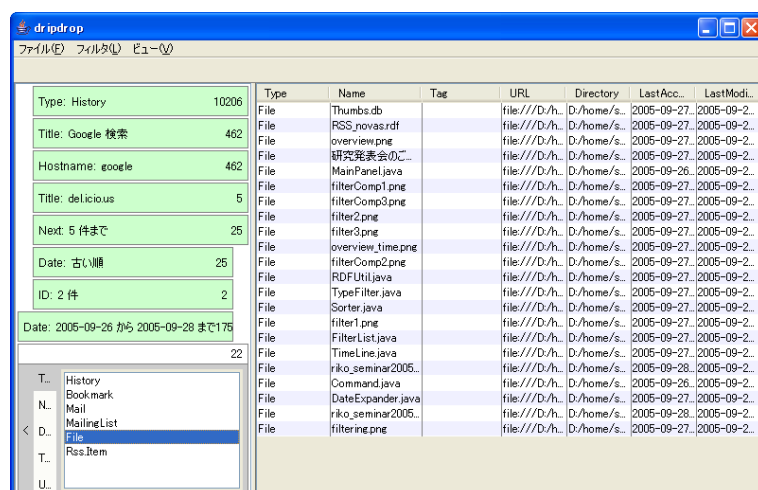


図 4.17 データの種類を制限 (b)

検索結果のファイルを眺めていると, riko_seminar.doc という名前のファイルを発見した (図 4.18) . そこでこの時期に理工学セミナーの発表があったことを思い出した.

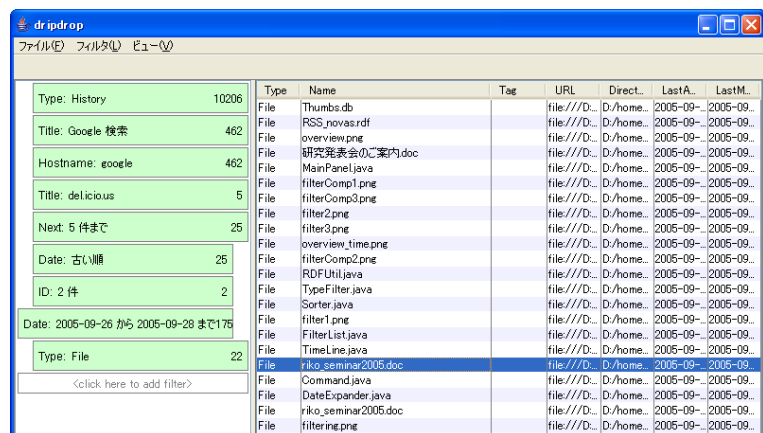


図 4.18 セミナーの資料を選択

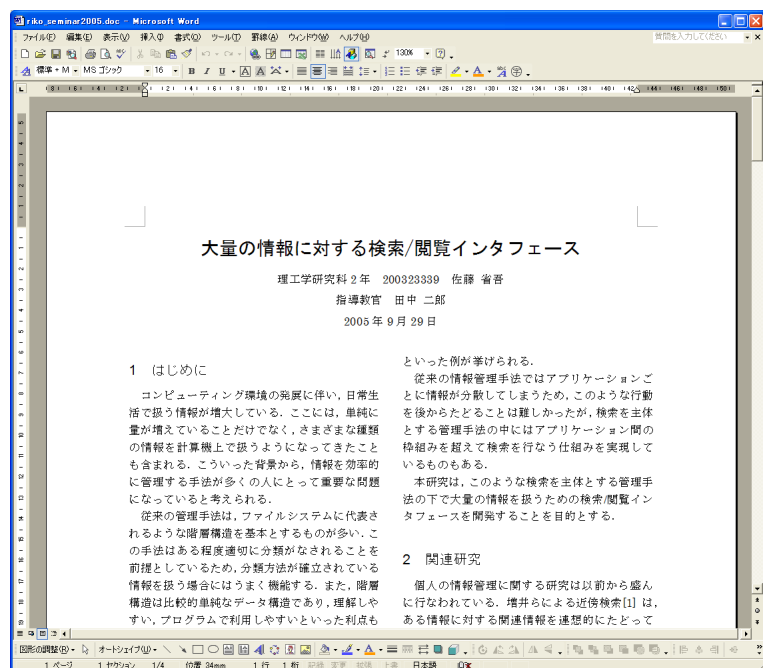


図 4.19 ドキュメントを開く

項目をダブルクリックしてドキュメントを開き、内容を確認した（図 4.19）。
以上で検索/閲覧の終了とする。

第5章 関連研究/関連システム

5.1 情報の管理と検索

分類によらない管理 分類ではなく、検索に重点を置いた管理手法が研究されている。増井らによる近傍検索 [11] は、ある情報に対して内容が近い・作成時刻が近い・格納場所が近いといった近傍の情報を連想的にたどることにより目的の情報を探し出す手法である。同じく増井による Q-Pocket[12] は、時刻に基づく ID のみで情報を管理するシステムである。また、暦本の提唱する Time-Machine Computing[13] は、時刻情報に注目して個人の情報管理を行い、任意の時刻の状態へと移動することを可能にする手法である。このような時刻に注目した情報管理としては超整理法 [14] がよく知られている。

また、Web 上に存在する Wiki サイトも分類によらないページ構造になっていると考えられる。通常の Web サイトはトップページをルートとする木構造に整理されていることが多いが、Wiki ではページ名を介して複雑にリンクしあったネットワーク構造になっている。

Semantic Web SIMILE Project[15] では、RDF を始めとする Semantic Web の技術を用いて、デジタル情報の相互運用性を高める研究を多数行っている。

Haystack[16] は、Eclipse のプラグインとして実装された情報管理アプリケーションである。RDF 形式でデータを一元管理しており、アイテムごとにビューやオペレーションが定義されている。また、PiggyBank[17] は Firefox の拡張機能として実装されており、Web から収集した RDF データを Web ブラウザ上で閲覧することができる。閲覧および検索のインタフェースは DHTML の技術に基づいており、同プロジェクトのブラウジングエンジンである Longwell[18] を使用している。Longwell の特徴は faceted browsing と呼ばれるインタフェースにある。facet とはデータ集合にとって重要だと考えられるメタデータのことであり、ユーザは facet に制限を加えていくことによってデータ集合の一部を取り出すことができる。

dripdrop のフィルタに基づく手法およびインタフェースは facet に基づくものとよく似ているが、dripdrop では、制限したデータ集合を別の観点から広げることに注目している点異なる。

デスクトップ検索 デスクトップ検索と総称される検索システムがいくつか現れている [19, 20, 21, 22, 23]。これらのシステムでは、計算機上のデータを常時監視してインデックスを作成することにより、手元のデータに対して Web 検索と同様の高速なキーワード検索を可能にしている。デスクトップ検索の大きな特徴として、アプリケーションを横断する検索ができ

るという点が挙げられる。この機能を OS レベルでサポートしているものもあり、今後は一般的な機能として広く普及していくと考えている。

その一方で、キーワードによって検索条件を指定する方式には問題が伴う。検索対象がいまいな場合や、適切なキーワードが思いつかない場合に対処できないという点である。本研究では、直接対象を指定できない状況に対し、近くのデータから集合を広げていくという手法を用いることでこの問題を解決しようとしている。

タグ付け 近年、タグを用いた分類手法が大きく注目されている。この手法では、ユーザは「タグ」と呼ばれるキーワードを個々のデータに付加することによりデータを分類し、後の検索に利用することができる。Web アプリケーションを中心とした多くのシステムで採用されている [1, 2, 24, 25, 3, 26]。なお、システムによっては同様の機能に別の名称を用いている場合もある。

従来の階層構造による分類と最も異なるのは、1つのデータに複数のタグを付けることができる点である。階層構造では同時に複数の親を持つことができないため分類の視点が固定されてしまうという問題があるが、タグ付けはこの問題点を解決した分類手法であるといえる。本研究においても、タグをメタデータの1つとして検索に利用している。

5.2 多次元属性情報の検索/閲覧

多次元の属性を持つデータに対する検索手法が提案されている。梶山らの Concentric Ring View F+[27] では、適合フィードバックを導入することによってスムーズな再検索を実現している。この手法は、ユーザがアイテムを選択した際にその属性値で検索を行うことにより、検索キーを調節することなく似たアイテムを探すことができるというものである。また、Lanning らの作成した MultiNav[28] では、注目したアイテムが常に中央に来るよう自動的に整列することで属性ごとにアイテム全体における位置付けを把握できる機能が実装されている。

両者に共通する特徴として、検索条件の設定と検索結果の表示を同一画面上で行っていることが挙げられる。dripdrop では検索に際し比較的詳細なパラメータ指定が可能であるが、その一方で、これらの画面が分かれていることにより操作の直感性が損なわれていると考えられる。

5.3 時系列に基づく視覚化

dripdrop で試作した時系列ビューでは円筒に巻きつけた螺旋のモデルを考えているが、平面上の螺旋によって時間の連続性と周期性を表現する手法は古くから見られる。例えば文献 [29] を参照すると、1888 年にそのような視覚化技術が用いられていることを確認できる。

最近では Carlis らによって螺旋を用いた視覚化手法が提案されている [30]。この研究では科学技術データの分析を対象としており、数十個のボタンを用いて表示方法を制御している。そのため日常の検索に使用できるような直感的なインタフェースであるとは言い難い。

また，リング上でポインティングデバイスを回転させる操作によってスクロールを実現する提案がなされている [31, 32].

第6章 議論

6.1 提案手法について

6.1.1 検索パラメータの指定

フィルタに基づく手法では、フィルタごとに詳細なパラメータを指定し、それを組み合わせることによって複雑な条件の検索を可能にしている。しかし、このような手法が理解しやすいかどうかについては疑問が残る。また、検索条件を明示的に指定しなければならないことから、ユーザに与える心理的負担が大きくなってしまうおそれがある。

関連研究および関連システムは Web ベースの実装になっているものが多い。この場合、条件の追加や類似データの取得といった操作は概ねハイパーリンクによって実現されている。日常の Web ブラウジングと同じようにクリックを繰り返すだけで希望するデータを取り出していく手法は、本研究の提案手法に比べてシンプルで分かりやすい側面があると考えられる。また、過去に行った操作はナビゲーションの履歴として参照できる。

一方、リンクによるナビゲーションではその一部分だけを取り出して他の状況に適用することが難しい。大量の情報に対して検索を行う場合は、よく使う条件をあらかじめストックしておき状況に応じて書き換えて使うことによって、ユーザが同じ条件指定を繰り返す手間を軽減することができるため、本研究で提案したような条件の再利用が容易な方式にも価値があるのではないかと考えている。

6.1.2 緩めるフィルタ

データ集合に対して条件を緩めるフィルタを適用する場合、現在の実装では、常にデータ集合全体を対象範囲としている。しかし実際の利用場面を考えてみると、メールのみを閲覧したい、今月のデータだけを見たいといったように緩める範囲を制限したい場合がある。すなわち、必ず満たさなければならない条件を指定することにより緩めるフィルタの及ぶ範囲を制御する機能があると便利だと思われる。

6.2 試作システムについて

6.2.1 データの取得

現在の実装では検索対象のデータを取得する機能がシステム内に組み込まれていないため、ユーザが自分でデータを用意する必要がある。そこで、さまざまな情報に関してメタデータを保持しているデスクトップ検索システムと連携を取ることができれば、実用性の向上が見込めると考えられる。本研究では一時期 Google Desktop API の利用を検討したが、キーワードによる検索が前提となっているため希望するデータがうまく得られず、利用を見送った。

また、プログラムによって自動的に生成されるメタデータとユーザが明示的に付加するメタデータを分けて扱うことにより、メタデータの重み付けを変化させる等の処理によって検索結果にユーザの意図をより反映させることができるのではないかと考えている。

6.2.2 インタフェース

フィルタを追加する際のマウス操作に注目すると、プロパティの選択や追加の確定はポインタの移動のみで行っている。一方、並べ替えやパラメータの設定ではボタンを使用して操作している。このように2種類の操作体系が混合しており、やや一貫性に欠ける部分がある。今後の検討が必要である。

時系列ビューについては、回転操作がマウスにはあまり向いていないという印象を得た。ペンまたはジェスチャのような入力手法や、ダイヤルのように回転操作のできるデバイスの方が提案したビューに適しているのではないかと考えている。

第7章 おわりに

本研究では，メタデータを利用して個人の情報を検索/閲覧するシステムを試作した．

このシステムでは，検索条件を構成する個々の操作をフィルタとして捉え，フィルタを繰り返し適用することにより段階的にデータを取り出していく手法を採っている．絞り込んだ結果に対して条件を緩めることにより，異なる観点から結果を拡張できることを特徴としている．また，時刻情報に注目してデータを閲覧する手法を提案し，ビューの試作を行った．

以上の提案手法を踏まえてシステム“dripdrop”を作成し，メタデータに基づく検索/閲覧を実現した．

謝辞

本研究は，多くの方々のご指導とご協力のもとに進められたものです。

指導教員である田中二郎先生や志築文太郎先生，また北陸先端科学技術大学院大学の三浦元喜先生には多くのご指導をいただきました。深く御礼を申し上げます。三末和男先生や高橋伸先生，ならびに田中研究室の皆様には，研究を進めるにあたって貴重な助言をいただきました。ありがとうございます。最後に，私を支えてくださった家族や友人に改めて感謝いたします。

参考文献

- [1] del.icio.us. <http://del.icio.us/>.
- [2] Flickr. <http://www.flickr.com/>.
- [3] Gmail. <http://mail.google.com/>.
- [4] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of ACM CHI'04*, pp. 415–422, 2004.
- [5] World Wide Web Consortium. <http://www.w3.org/>.
- [6] Resource Description Framework (RDF). <http://www.w3.org/RDF/>.
- [7] RDF Site Summary (RSS) 1.0. <http://web.resource.org/rss/1.0/>.
- [8] the friend of a friend (foaf) project. <http://www.foaf-project.org/>.
- [9] Jena. <http://jena.sourceforge.net/>.
- [10] Piccolo. <http://www.cs.umd.edu/hcil/piccolo/>.
- [11] 増井俊之, 塚田浩二, 高林哲. 近傍関係にもとづく情報検索システム. 第11回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2003) 論文集, pp. 79–86, 2003.
- [12] 増井俊之. 検索と例情報を活用した情報管理手法 Q-Pocket. インタラクティブシステムとソフトウェア VIII, pp. 191–196, 2000.
- [13] Jun Rekimoto. Time-Machine Computing: A Time-centric Approach for the Information Environment. In *Proceedings of ACM UIST'99*, pp. 45–54, 1999.
- [14] 野口悠紀雄. 「超」整理法 1–4. 中公文庫. 中央公論新社, 2003.
- [15] SIMILE Project. <http://simile.mit.edu/>.
- [16] David Karger, Karun Bakshi, David Huynh, Dennis Quan, and Vineet Sinha. Haystack: A General-Purpose Information Management Tool for End Users Based on Semistructured Data. In *Proceedings of the 2005 CIDR Conference*, pp. 13–27, 2005.

- [17] David Huynh, Stefano Mazzocchi, and David Karger. Piggy Bank: Experience the Semantic Web Inside Your Web Browser. In *Proceedings of International Semantic Web Conference (ISWC) 2005*, 2005.
- [18] SIMILE | Longwell. <http://simile.mit.edu/longwell/>.
- [19] Google Desktop. <http://desktop.google.com/>.
- [20] Windows Desktop Search. <http://desktop.msn.com/>.
- [21] Yahoo! Desktop Search. <http://desktop.yahoo.com/>.
- [22] Apple - Mac OS X - Spotlight. <http://www.apple.com/macosx/features/spotlight/>.
- [23] Main Page - Beagle. <http://beaglewiki.org/>.
- [24] はてなブックマーク. <http://b.hatena.ne.jp/>.
- [25] CiteULike. <http://www.citeulike.org/>.
- [26] Picasa. <http://picasa.google.com/>.
- [27] 梶山朋子, 神門典子, 佐藤真一. Concentric Ring View F+: リング状検索 GUI における履歴と適合フィードバック. *インタラクション 2005 論文集*, pp. 197–198, 2005.
- [28] Tom Lanning, Kent Wittenburg, Michael Heinrichs, Christina Fyock, and Glenn Li. Multidimensional Information Visualization through Sliding Rods. In *Proceedings of ACM AVI 2000*, pp. 173–180, 2000.
- [29] Edward Rolf Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [30] John V. Carlis and Joseph A. Konstan. Interactive Visualization of Serial Periodic Data. In *Proceedings of ACM UIST'98*, pp. 29–38, 1998.
- [31] Tomer Moscovich and John F. Hughes. Navigating Documents with the Virtual Scroll Ring. In *Proceedings of ACM UIST'04*, pp. 57–60, 2004.
- [32] G.M. Smith and m.c. schraefel. The Radial Scroll Tool: Scrolling Support for Stylus- or Touch-Based Document Navigation. In *Proceedings of ACM UIST'04*, pp. 53–56, 2004.