修 士 論 文

京大型カードに記録したデータを有効に利用できる発想支援ツールの構築

指導教官　國藤 進 教授

北陸先端科学技術大学院大学
知識科学研究科知識社会システム学専攻

050021　金丸 浩士

審査委員:　國藤 進 教授 (主査)
藤波 努 助教授
西本 一志 助教授
伊藤 孝行 助教授

2002 年 2 月

目 次

図 目 次

表 目 次

第 1 章

序論

1.1 本研究の背景

近年さまざまな分野での電子化が進み，オフィスや学校に急速に PC が普及している．また，発想支援ツールの研究も盛んに行われている．

國藤のモデルによれば，人間の創造的問題解決のプロセスは発散的思考，収束的思考，アイディア結晶化，評価・検証の 4 つの段階から構成される．発散的思考とは提起された問題を明らかにし，提起された問題に対して情報収集し現状分析を行なう段階である．収束的思考は，発散的思考で得られた情報を元に問題解決のための本質的情報を読みとっていく段階である．このプロセスの中心は仮説の生成である．この段階を経て複数の仮説が生成されたあと，アイディア結晶化の段階で仮説を評価し，問題解決に最も有効と評価される仮説を直観的に評価し，採択する．最終的に採択された仮説を，実際に現実世界で実行し，仮説が正しかったかどうかをその結果により検証するのが評価・検証の段階である [3].

これらのプロセスを支援するためのさまざまな手法が存在する．

ブレーンストーミングは Osborn 代表的な発散的思考技法である．ブレーンストーミングは

1. 批判禁止
2. 自由奔放の歓迎
3. 質より量を求める

4. 他人のアイディアへの便乗の歓迎

という4つの原則の元，参加者が思いつく限りの発言を行なう，という形式で行なわれる．

KJ法 [1] は川喜田二郎により提唱された，問題解決技法である．事実，推定，意見などを統合することによって，対立する意見，期待を裏切る結果，現実との矛盾の中から，問題の本質などを明確化するための手法である．KJ 法の名は考案者の川喜田二郎に由来している．具体的な手順として

1. データの収集とデータ整理
2. 一行見出しの作成とそれを元にしたブレーンストーミング
3. ラベルのグルーピング
4. 表札作り
5. ラベル群配置の決定
6. 図の作成 (A 型図解化)
7. シナリオ作成 (B ’型口頭発表もしくは B 型文章化)

という順番で行なう．

京大型カードは梅棹忠夫が文献 [2] で示した B6 判の大きさの紙に罫線が引いたカードである．京大型カードはデータ整理の道具として利用される．1 枚のカードに 1 つの情報を書き込む．また書き込む情報として，タイトル (一行見出し)，本文，記入した日付，記入した場所，記入者，情報源がある．

1.2 関連研究

上記の作業を支援するような，さまざまなコンピュータを利用した発想支援システムの研究が行なわれている．

発想支援グループウェア郡元 [4] は，KJ 法をモデルにした発想支援システムである．郡元は京大型カードシステムを参考にして作成されたカード型データベースである知的生産支援システム Wadaman を備え，フィールドワークの結果や KJ 法の結果などの保存

を行い，KJ 法作業の支援ツールとなっている．また PDA 端末を用いて手書きインターフェースを用いることによりいつでもどこでもアイディアや情報をメモすることにより，KJ 法作業にフィールドワークで集めたデータを生かすことができる．本システムは，ベクトル空間モデルを用いた発想支援機能を有する．

金子は発散的思考の時に発想しているテーマに関連するホームページからテキストマイニング技術を利用してユーザのすでに出しているアイディアをもとにヒントを与える発散的思考支援システムを構築した．本システムでは，ヒントとして得られる情報が文字データだけではなく画像も利用することができる [13].

金井はベクトル空間モデルを利用した情報フィルタリング技術を利用し，会議における場の文脈を元に対話場を活性化することを試みている [11]．本システムではデータを XML で管理しているため，データの受渡しの必要が生じた場合，スムーズに行なうことができる．

1.3 本研究の目的

通常の KJ 法作業では，現状把握ラウンドにおいて，フィールドワークで集めてきたデータをカードに整理し，一行見出しをつくる．そして一行見出しを元にしてブレーンストーミングを行い，ラベル化する．A 型図解にまとめ，B ’型口頭発表もしくは B 型文章化し，結果をまとめる．その際にフィールドワークによって得たデータと当人の記憶を頼りにカードを作成するため、見聞きしたすべての事実を反映できるとは限らない．

本研究ではこの問題を解決するために京大型カードに記入したデータを有効に利用できるような発想支援ツールを構築する．

1.4 本論文の構成

本論文は 7 章で構成される．2 章では，本システムの概要について述べる．3 章では，本システムを構築するに当たり利用した技術について述べる．4 章では，本システムの動作の仕組みについて述べる．5 章ではインタフェース部分の実装について述べる．6 章では本システムの評価実験の結果について述べ，結果に対し考察を与える．7 章では，本

研究で得られた成果と今後の課題について述べる.

第 2 章

システムの概要

本章では，京大型カードに記録したデータを有効に利用できる発想支援ツールの概要について説明する．

本システムは大きく分けて以下の 2 つのモジュールに分けられる

1. 索引付け、転置ファイル作成モジュール

2. 文書問い合わせモジュール

全体モジュール図を 2.1 に示す．

作業工程としては，まず，XML ファイルにフィールドワーク等で集めてきた文字データ，画像データを登録する．その次に索引付け、転置ファイル作成モジュールで索引付けをする．索引付けとは その文書の内容を良く表していると考えられる要素を抽出し，その集合で文書の内容を表現する方法である．その後転置ファイルに索引付けの内容を書き出す．転置ファイルとは，索引語とそれに対する位置リストからなる．位置リストとは，その索引語があらわれている文書名を列挙したリストのことである．

そして，ブレーンストーミング作業中等にアイディアのヒントを得るために，問い合わせキーワードを文書問い合わせモジュールに送信する．文書問い合わせモジュールは問い合わせキーワードと文書の類似度を計算し，文書が類似している順番に結果を出力する．

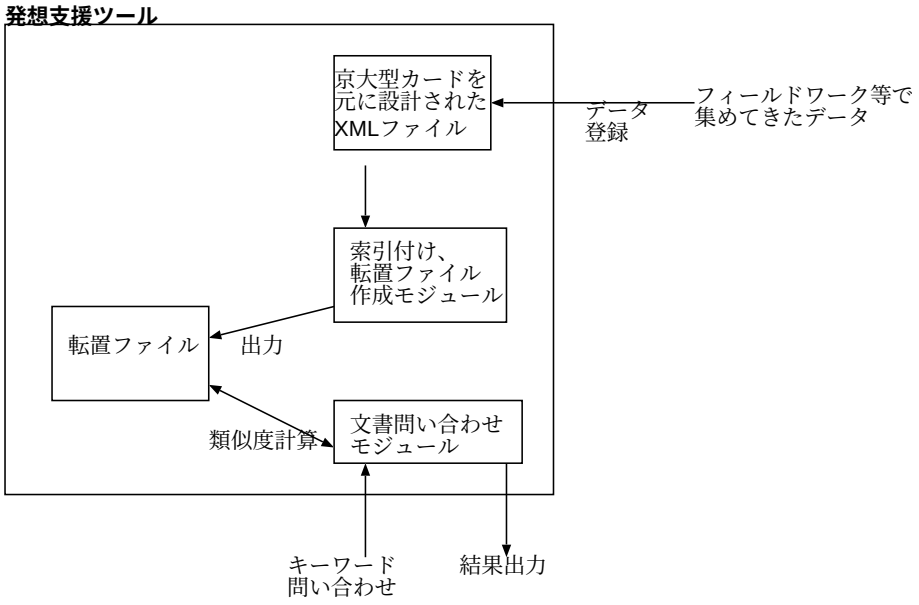

図 2.1: 全体モジュール図

2.1 システムの機能

本システムの機能は主に 3 つあげられる．

データ保存機能　フィールドワークなどで集めてきたデータを京大型カードを元にして設計した XML ファイルに登録し，保存することができる．保存できるデータはタイトル，本文，日付，記録者，場所，画像ファイル名である．

キーワードによる文書問い合わせ機能　KJ 法におけるブレーンストーミング作業時にベクトル空間モデルを利用して入力したキーワードと京大型カードとの類似度を計算し類似度が高い順に出力される．

京大型カードによる文書問い合わせ機能　上記作業時に類似の京大型カードを探したくなった時は京大型カードのファイル名を指定して，上記と同様の文書の問い合わせを行なう．

第 3 章

発想支援ツールの利用技術

本章では，ツールを作成するにあたり利用した技術について説明する．

3.1 XML の利用

XML(eXtensible Markup Language) [6] とは Web 上で構造化文書をやりとりするためのデータフォーマットである．W3C(World Wide Web Consortium) により基本仕様を策定された．

特徴は，ユーザが独自のタグを使ってデータの属性情報や論理構造を独自に定義できることである．データの属性とデータのないようを関連づけて記述できる．

本システムでは，京大型カードに基づいて XML の構造を設計した．

3.2 文書への索引づけ

今回行った処理は以下に挙げるものである．

1. 不要語の削除
2. 出現頻度に基づく索引語の重みづけ

以下では処理過程について説明をする．

3.2.1 不要語の削除

不要語とは，日本語であれば「これ」「する」など，英語であれば “a”, “the”, “this”, “when”, などのように，文書の内容を特徴づけるのにほとんど役に立たない語のことを指す.

通常このアプローチで一般的に用いられるのは不要語のリストである不要語リストをあらかじめ作っておき，その中にあらかじめ含まれている語を削除するという手法である．しかし本システムでは文書を形態素解析し，文書を形態素に分け，各形態素に品詞情報を付与する．そして，名詞，動詞，形容詞のみを索引語として採用した．助詞や助動詞など文書の内容を特徴づけるのにほとんど役に立たない語は削除した.

3.2.2 索引語の重みづけ

索引語の重みを利用することで，各索引語の文書中や文書集合中での重要度を考慮して問い合わせキーワードに対するに対する文書の適合度を反映し，結果表示をさせる.

索引語の重みづけにおいてしばしば用いられる代表的なアプローチとして，TF・IDF 重みづけ (TF・IDF weighting) がある [9]．本システムでは索引語の重みづけ手法に TF・IDF 重みづけを採用している．以下ではこの手法について説明する.

TF(term frequency)

TF・IDF 重みづけの基礎のひとつとなるのが TF である．これは，ある文書 d 中に出現する索引語 t の頻度のことであり，$tf(d,t)$ で表す.

$tf(d,t)$ を単純に求めるアプローチとしては出現回数 $f(d,t)$ を用いて

$$tf(d,t) = f(d,t) \tag{3.1}$$

と求める手法がある.

しかし，このような計算方法だと文書の長さが長いほど索引語頻度がおおきくなってしまうため，ふたつの文書間で索引語の頻度をくらべたりする場合に都合が悪いという欠点がある．そこでしばしばとられるのが，その文書中の索引語の出現頻度の総和を計

算し，その総和で各索引語の出現頻度を割ることで，文書の長さによらない相対的な頻度として索引語頻度を与える手法である． [9] では以下のような式が示されている．

$$tf(d,t) = \frac{f(d,t)}{\sum_{s \in d} f(d,t)} \tag{3.2}$$

また索引語頻度はこのような式に限らず，たとえば， [10] では

$$tf(d,t) = 1 + \log f(d,t) \tag{3.3}$$

という式が示されている．(3.3) 式では索引語の対数をとることで，出現回数が索引語に与える影響を控えめにしている．(3.2) 式にくらべると，文書全体の索引語の総数を求めないため，計算コストの面で若干有利である．計算コストを減らすことにより文書問い合わせの結果出力が早くなり，ユーザの思考の流れを止める可能性を減らすことができる．よって本システムでは (3.3) 式を利用している．

IDF(inverse document frequency)

IDF はある索引語が全文書中のどれくらいの索引語が全文書中のどのくらいの文書に出現するかを表す尺度である． [9] や [10] では

$$idf(t) = \log\left(\frac{N}{df(t)}\right) + 1 \tag{3.4}$$

という式が示されている．ここで N は文書集合中の文書の総数であり，$df(t)$ は索引語 t が出現する文書数である．特にある索引語 t がどの文書にもあらわれる場合， $idf(t) = \log(N/N) + 1 = 0 + 1 = 1$ となり，これが IDF の最小値となる．対数をとる理由は，文書集合に対して IDF の値の変化をちいさくするためであり，一般には対数をとった方がよい問い合わせの結果が得られるとされる．本システムでは (3.4) 式を利用している．

TF・IDF 重み付け

TF・IDF 重み付けとは，上記の TF と IDF を組み合わせたものである．ある文書 d における索引語 t を TF・IDF 重み付けに基づいて重みづけをすると，索引語の重み $w(d,t)$ は，

$$
\begin{aligned}
w(d,t) &= tf(d,t) \cdot idf(t) && (3.5)\\
&= (1+\log f(d,t)) \cdot \left(\log\left(\frac{N}{df(t)}\right)+1\right) && (3.6)
\end{aligned}
$$

と与えられる．TF・IDF 重みづけのねらいはある文書内のおける高頻度語の重みづけを高くし，文書全体における高頻度語の重みづけを低くすることにある．

3.3 ベクトル空間モデル

本システムでは文書問い合わせの手法としてベクトル空間モデルを利用する．ベクトル空間モデルは，各文書をベクトルととらえるところからその名前がある．たとえば文書集合全体で T 個の索引語が得られたとする．このとき各文書を T 次元のベクトルで表すことができる．各ベクトルの要素は，その次元に対応する索引語をその文書が何個含んでいるかという頻度である．ある文書 n のキーワードベクトルは $\vec{d_n} = (t_1, t_2, \cdots, t_T)$ と表現できる．

ここで，例として以下のようなキーワードベクトルを考える．

$$
\begin{aligned}
\vec{d_1} &= (1,2,3)\\
\vec{d_2} &= (4,0,0)\\
\vec{d_3} &= (2,0,5)
\end{aligned}
$$

ただし，文書数は 3，T=3 としている．

上のようなキーワードベクトルが得られたとき，ベクトル間の類似度を求めるときに余弦尺度 (cosine measure) を用いる．$\vec{d_x} = (x_1, x_2, ..., x_T)$, $\vec{d_y} = (y_1, y_2, ..., y_T)$ を 2 つのキーワードベクトルとしたとき，余弦尺度に基づく類似度 $sim(\vec{d_x}, \vec{d_y})$ は以下のようになる．

$$
sim(\vec{d_x}, \vec{d_y}) = \frac{\sum_{i=1}^{T} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{T} x_i^2}\sqrt{\sum_{i=1}^{T} y_i^2}} \qquad (3.7)
$$

これは，ベクトルの要素ごとの積をとった和を分子とし，分母にベクトルの大きさの積をとったものである．また (3.7) 式は，

$$sim(\vec{d_x}, \vec{d_y}) = \frac{\vec{d_x} \cdot \vec{d_y}}{|\vec{d_x}||\vec{d_y}|} \tag{3.8}$$

と書くことができる．ただし，$\vec{d_x} \cdot \vec{d_y}$ は $\vec{d_x}$ の内積を示し，$|\vec{d_x}||\vec{d_y}|$ は $\vec{d_x}$ と $\vec{d_y}$ の大きさの積を示す．これはベクトルの角を θ としたときに $\cos\theta$ を求める計算である．

ベクトル空間モデルでは全てキーワードベクトルの要素が非負であるので，2 つのキーワードベクトルのなす角度の範囲は $0 \leq \theta \leq \pi/2$ である．類似度の値は，最も大きくなるのが $\theta = 0$ のときで $\cos\theta = 1$，もっとも小さくなるのが $\theta = \pi/2$ のときで $\cos\theta = 0$ である．また $\theta = 0$ ということはキーワードベクトルがおなじ方向を指しているということを示す．$\cos\theta$ が大きいほど類似しているということを示す．

余弦尺度に基づいて，上の例における $\vec{d_1}$ とほかの 2 つのキーワードベクトル間の類似度を求めると以下のようになる．

$$\begin{aligned} sim(\vec{d_1}, \vec{d_2}) &= \frac{1 \times 4 + 2 \times 0 + 3 \times 0}{\sqrt{1^2 + 2^2 + 3^2}\sqrt{4^2 + 0^2 + 0^2}} = 0.27 \\ sim(\vec{d_1}, \vec{d_3}) &= \frac{1 \times 2 + 2 \times 0 + 3 \times 5}{\sqrt{1^2 + 2^2 + 3^2}\sqrt{2^2 + 0^2 + 5^2}} = 0.84 \end{aligned}$$

よって文書 $\vec{d_1}$ に似ているのは $\vec{d_3}$,$\vec{d_2}$ の順になる．

上記の例は文書間の類似度の計算であったが，ユーザから与えられる問い合わせキーワードと文書の間のことであっても同様の処理となる．

3.4 転置ファイル

通常，文書問い合わせの手法にベクトル空間モデルを用いた場合，文書数と索引語に比例して計算量が増大する．それにより，問い合わせキーワードと文書間のキーワードベクトルの余弦尺度の値を計算する処理に時間を要してしまい，思考の流れを止めてしまうことになりかねない．本システムではこの問題を軽減するために転置ファイルを用いている．

転置ファイル (inverted file) とは索引語をキー値として探策可能な索引のことである．3.3 節で 3 件の文書のキーワードベクトル例を示したが，この例に対する転置ファイルは表 3.1 のようになる．

表 3.1: 転置ファイルの構成

索引語	位置リスト
t_1	$\langle d_1, 1\rangle, \langle d_2, 2\rangle, \langle d_3, 3\rangle$
t_2	$\langle d_1, 4\rangle$
t_3	$\langle d_1, 2\rangle, \langle d_3, 5\rangle$

転置ファイルは，索引語とそれに対する位置リストからなる．位置リストとは，その索引語があらわれている文書名を列挙したリストのことである．

転置ファイルの役割は，与えられた索引語に対し，対応する位置リストを瞬時に得ることができるようにすることである．本システムでは転置ファイルを利用することにより，問い合わせキーワードに用いられている単語が含まれていない文書との類似度計算と，内積を計算するときにそれぞれの計算結果が 0 になる計算を省略することができる．

第 4 章

発想支援ツールのしくみ

本章ではXMLファイルで設計した京大型カードと発想支援システムにおける，索引付け，キーワード問い合わせの仕組みについて説明する．

4.1 京大型カードをもとにしたXMLファイル

京大型カードをもとにXMLファイルを以下に示すようにデザインした．京大型カードの構造を図4.1に示す．

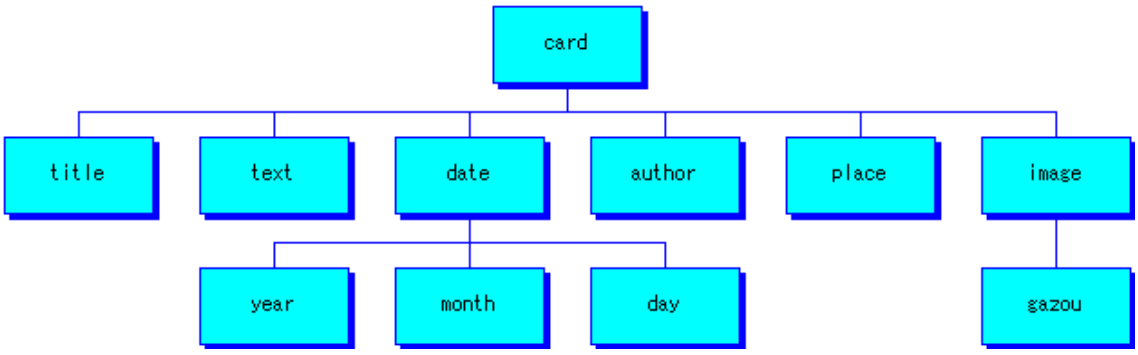

図 4.1: XML 階層図

XMLではデータ構造がタグの入れ子により階層として表現される．
例えば図4.1の例だと

<card>

<title></title>
<text ></text >
<date>
<year></year>
<month></month>
<day></day>
</date>
…(中略)
</card>

という構造になる．

XML 文書で設計した京大型カードのタグの意味を以下に説明する．

card：全てのルートとなる部分

title：題名を入力する

text：本文を入力する

date：データカードに記入した日時を記入する．記入する内容は year，month，day にそれぞれ年，月，日を記入する．

author：データ入力者の氏名を記入する．

place：データを入力した場所を記入する．

image：画像データを記入する．image はそのタグに数字をつけることにより複数の image タグの出現を許している．これにより複数個の画像データの登録を実現させている．gazou タグ内には画像のファイル名を記入する．

4.2 索引付け、転置ファイル作成

本システムではキーワード問い合わせ速度の高速化のために転置ファイルを用いている．従ってそのため前処理が必要となる．

作業の流れを図 4.2 に示す．索引作成の流れは以下のとおりである．

1. 文書に対し文書 ID をつける．XML 文書のファイル名を読みこんでいって，読みこんだ順に 1，2，3，… と番号をつけていく．これは転置ファイルに文書名を書

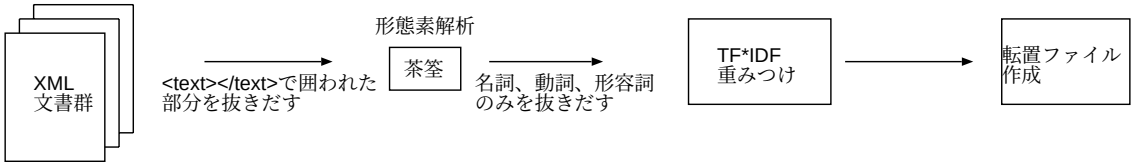

図 4.2: 索引づけ

き込むのは無駄が大きいため各文書に文書 ID をつけ文書名の代わりに利用するために行う.

2. 京大型カード XML ファイルの <text></text>と囲まれた部分を抜きだす. 次に抜き出した文章を茶筌 [5] で形態素解析し，形態素解析の結果，名詞，動詞，形容詞であるもののみを抜き出し，索引語とする..

3. TF・IDF を計算し，キーワードベクトルを構築する.

4. 文書問い合わせ時の計算量軽減のため，転置ファイルや補助的なデータを作成する. ここで作成するのは，文書名と文書 ID の対応付を記したファイル，文書 ID とキーワードベクトルの長さを記したファイル，文書 ID に関する転置ファイル，TF 値に関する転置ファイル，IDF 値に関する転置ファイル，TF・IDF に関する転置ファイルである.

4.3 類似文書のキーワードを利用した問い合わせ

ここではベクトル空間モデルによるキーワードを用いた問い合わせについて説明する. 作業の流れを 4.3 に示す.

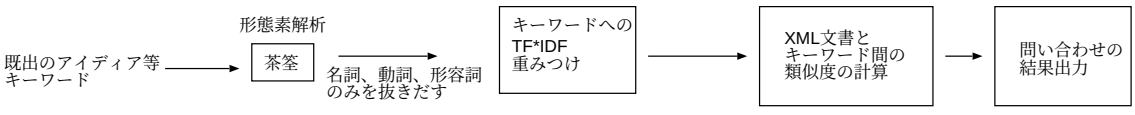

図 4.3: キーワードを利用した問い合わせしくみ

キーワード問い合わせにおけるアルゴリズムは以下のとおりである.

1. 問い合わせのキーワードを形態素解析し，名詞，動詞，形容詞であるもののみを抜き出し，索引語とする．

2. 索引語に対し重みづけをする．問い合わせキーワード q の各要素の重みは

$$w(q,t) = tf(q,t) \cdot idf(t) \tag{4.1}$$

   となる．ただしこのときの IDF は計算の簡単化のため，索引作成の段階で求めた値を用いている．

3. さきほど構築した転置ファイルより文書長や各々の文書の索引語への重みつけされた値を取り出し検索キーワードと文書のキーワードベクトルの余弦尺度を計算する。

4. そして各文書との類似度を計算し，類似度順にソートする．

以上がキーワードによる文書問い合わせの仕組みである

後述するが，本システムはキーワード問い合わせ機能のほかにある文書を指定することによる類似文書の問い合わせ機能を有している．この機能では指定された文書の<text></text>で囲われた部分を問い合わせキーワードとしている．問い合わせの原理はキーワードを用いたものと同様である．

第 5 章

ユーザインターフェース

本章では，発想支援ツールのインタフェースについて説明する．本システムの初期画面を図 5.1 に示す.

図 5.1: 初期画面

インタフェースには Web ブラウザを用いており，CGI によってキーワード問い合わせの処理を行っている．またキーワード問い合わせにおける処理言語は Perl を用いている.

5.1　キーワードによる問い合わせ

ここではキーワード入力による問い合わせについて説明する．発散的思考作業時にアイディアが出なくなったら，今まで出ているアイディアやその他のキーワードを入力語とし，ヒントを得るなどの利用法が考えられる．利用画面を図 5.2 示す.

図 5.2: キーワード問い合わせ結果一覧

ブラウザのフレーム左側の,「キーワードから問い合わせる」のほうにマークをし，問い合わせキーワード入力後,「クエリ送信」ボタンを押すとブラウザのフレーム右側に問い合わせの結果が出力される.

5.2　京大型カードを利用した問い合わせ

キーワード問い合わせで見つけた京大型カードのデータに似ているカードをさがしたい時は，京大型カードを利用した問い合わせ機能を利用する。利用画面を図 5.3 に示す.

操作方法はで説明した内容とほぼ同様である．ブラウザのフレーム左側の,「XML ファイルから選択する」のほうにマークをし，ファイル名を入力後,「クエリ送信」ボタンを押すとブラウザのフレーム右側に問い合わせの結果が出力される.

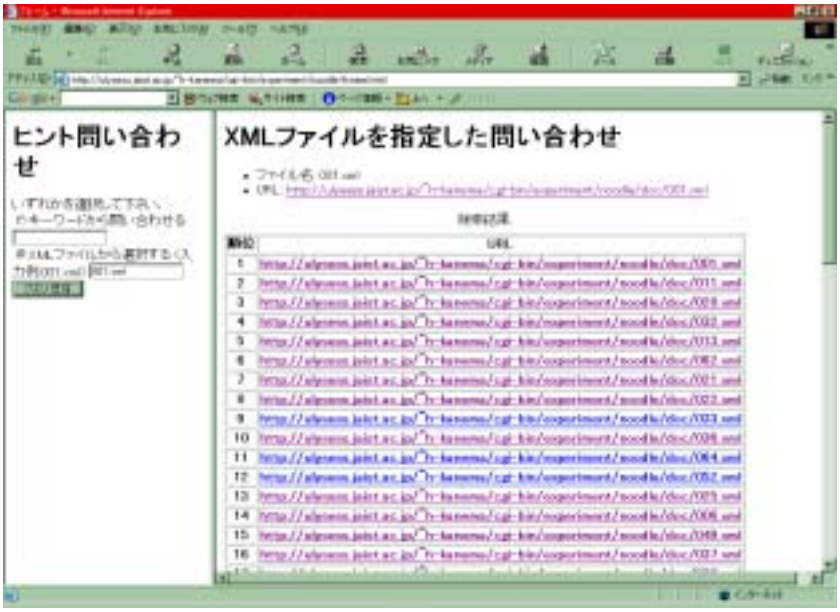

図 5.3: 類似文書問い合わせ結果一覧

5.3 京大型カードの出力

5.1 節や 5.2 節で説明した結果一覧で，URL をクリックすると図 5.4 のような京大型カードに基づいたデータカードが出力される．また XSL(eXtensible Stylesheet Language) [7] を利用し，XML を HTML に変換しブラウザへの表示させている．

図 5.4: 京大型カードの出力

第 6 章

評価実験

6.1 実験の目的

本システムはKJ法作業におけるブレーンストーミング作業の支援を目的としている．本システムの特徴は文書データだけではなく画像も提示できることが特徴である．今回の実験の目的は画像データが発想に有効であるかどうかを評価することにある．

6.2 実験方法

発想支援システムの評価方法として，新しい機能を用いたときと用いなかったときについての比較を行うことが多いが，今回は検索エンジン Google [12] と本システムの比較を行った．通常インターネットで情報を集めるときに使うのは検索エンジンを利用するため，今回比較の対象として Google を用いた．

被験者は北陸先端科学技術大学院大学 知識科学研究科の学生 20 名対象に行った．

作業は被験者 1 人ずつに作業してもらった．KJ 法作業に則って評価を行った．作業は以下に示す順番で行なった．

1. ブレーンストーミングを 20 分間
2. 表札作り
3. 衆目評価

実験に用いられたテーマは以下の 2 題である.

テーマ 1： あたらしいラーメンのアイディアを出す

テーマ 2： 環境問題を解決するには

各被験者に行なってもらったテーマを表 6.1 に示す.

表 6.1: 被験者の課題

	実験に利用するツール	
	google	発想支援ツール
a 群	テーマ 2	テーマ 1
b 群	テーマ 1	テーマ 2

被験者を a 群と b 群それぞれ 10 名づつにわけた．a 群の被験者 a1〜a10 には課題 2 をまず Google[12] を利用してブレーンストーミング作業をしてもらった．KJ 法作業が一通り終ったら，次に課題 1 を本システムを利用して，評価した．逆に b 群の被験者 b1〜b10 には課題 1 をまず Google を利用してブレーンストーミング作業をしてもらい，次にテーマ 2 を本システムを利用して，評価した.

本システムの実験時に被験者が参照するデータとして，a 群の被験者にはラーメン店を紹介した書籍から 65 ページ分，b 群の被験者にはリサイクルエネルギーについて書かれた書籍から 66 ページ分を提示した.

定性的評価として発想支援ツールに関するアンケートをとった．アンケートの項目は以下の通りである.

質問 1：あなたは KJ 法やブレーンストーミングをやったことがありますか? (はい/いいえから選択)

質問 2：発想支援ツール全体の操作感は良かったですか?悪かったですか? (良かった/悪かったから 5 段階で選択) またその理由もおねがいします.

質問 3：発想支援ツールの結果の表示方法は良かったですか?悪かったですか?(良かった/悪かったから 5 段階で選択) またその理由もおねがいします.

質問 4：発想支援ツールの文書をキーワード問い合わせできる機能は良かったですか?悪かったですか?(良かった/悪かったから 5 段階で選択) またその理由もおねがいします.

質問 5：発想支援ツールの類似文書を検索できる機能は良かったですか?悪かったですか?(良かった/悪かったから 5 段階で選択) またその理由もおねがいします.

質問 6：発想支援ツールにより示された文書データはラベルの作成に役に立ちましたか?役に立ちませんでしたか?(役に立った/役に立たなかったから 5 段階で選択) またその理由もおねがいします.

質問 7：発想支援ツールにより示された画像データはラベルの作成に役に立ちましたか?役に立ちませんでしたか?(役に立った/役に立たなかったから 5 段階で選択) またその理由もおねがいします.

質問 8：発想支援ツールを利用することにより満足がいくアイディアが出ましたか?出ませんでしたか?(出た/出なかったから 5 段階で選択) またその理由もおねがいします.

質問 9：そのほか気づいたことがありましたらご自由にお書き下さい.

定量的評価として画像が発散的思考に有効であると答えたユーザの b 群におけるテーマ 1 でのラベル数の合計と a 群におけるテーマ 1 でのラベル数の合計を比較した．またテーマ 2 に関しても同様の比較をした.

6.3 実験環境

実験に使用した環境は以下のとおりである.

6.3.1 サーバ側

OS: Vine Linux 2.1.5(Calon-Segur)

カーネル: 2.2.18-0vl4.2

CPU: Intel Pentium II 400MHz

メモリ: 128MB

HDD: 6GB

Web サーバ: Apache 1.3.20

CGI スクリプト: Perl (jperl 5.005_03)

形態素解析アプリケーション: 茶筌 version 2.2.8

6.3.2 クライアント側

OS: Windows 95 Second Edition

CPU: Pentium III 750MHz

メモリ: 128MB

Web ブラウザ: Internet Explorer 5.0

KJ 法ソフト: 電子 KJ 法

KJ 法作業を行うアプリケーションとして，川喜田研究所の電子 KJ 法を利用した．電子 KJ 法の利用画面を図 6.1 に示す．

6.4 結果

6.4.1 定性的評価

本システムの画像データについて尋ねた質問 7 に関する評価と理由を表 6.2 に示す．また質問 7 に関する評価をグラフにしたものを図 6.2 に示す．またアンケートの全結果は付録 A に示してある．

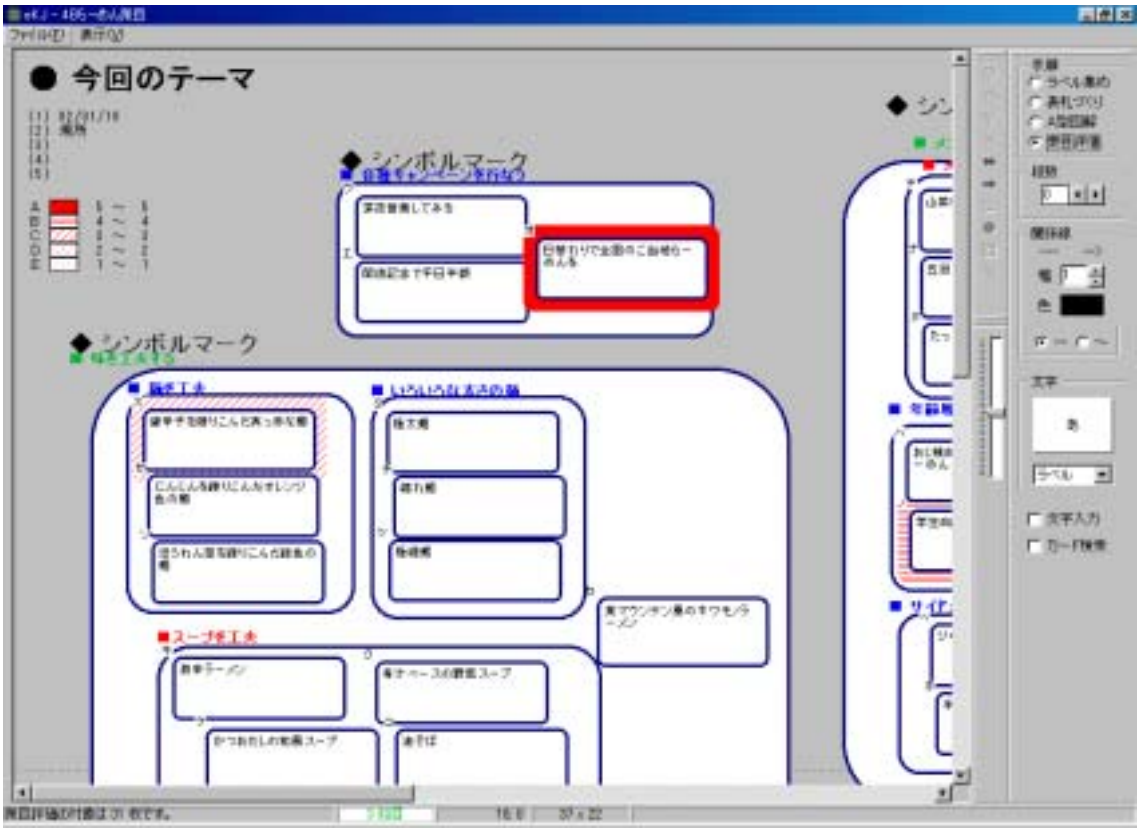

図 6.1: 電子 KJ 法の作業画面

表 6.2: 質問 7 の点数と理由

被験者	点数	理由
a1	5	ラーメンの絵を見てピンとひらめいた
a2	4	画像で文章にはないイメージを想起できた
a3	5	
a4	4	味覚を刺激されたので
a5	1	画像より文書に注目していました
a6	1	ラーメン画像からの発想が得られず
a7	2	画像より文章の方役立った．文章の方が具体的で分かりやすかった.
a8	4	今回のような場面ではあまり役に立たなかったが, ふだん使用するときには画像イメージがあるのは大きいと思います
a9	5	
a10	2	文書を見る方に集中していたのであまり役に立たなかった.
b1	4	アイデアを出す助けになった
b2	2	見にくかったです
b3	2	ラベルが文字で作られるため
b4	4	画像が見やすかったので
b5	4	文章だけより図を見て理解しようとした
b6	2	図が意図することがわかりにくかった
b7	5	文より画像のほうが見やすい
b8	3	視覚でとらえられる
b9	1	ほとんど見てないから
b10	2	下部に隠れるため見えない
平均	3.1	

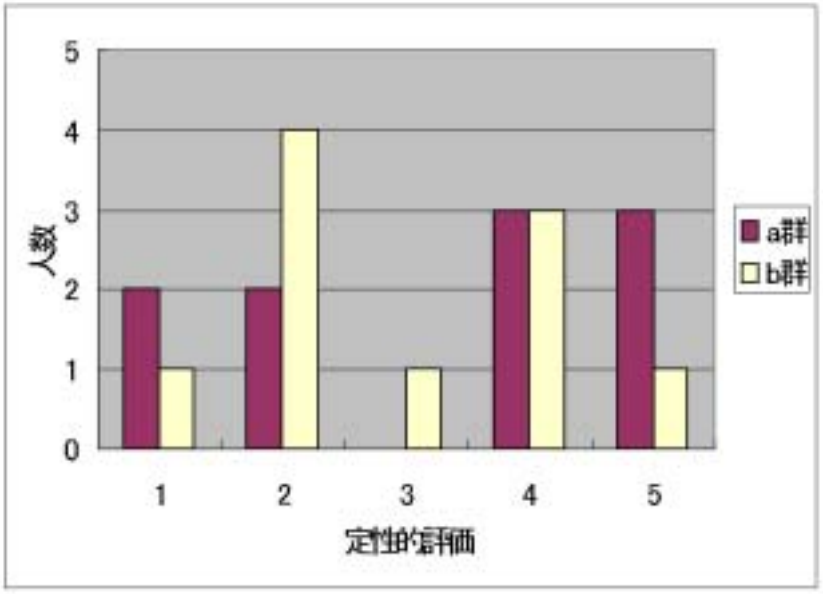

図 6.2: Q7 の結果

6.4.2 定量的評価

発散的思考作業に画像が役に立ったと言う被験者のアイディア数を集計したものを表 6.3 と図 6.3 に示す．表 6.3 は被験者 a1, a2, a3, a4, a9, b1, b4, b5, b7, b8 のラベル数の平均値である．また各被験者のラベル数は付録 B に示してある．

表 6.3: 画像が発想に役に立ったという被験者のラベル数

	実験に使用したツール	
テーマ	Google	発想支援ツール
テーマ 1	24.2	25.2
テーマ 2	22.6	22.8

また Google での実験におけるラベルの枚数でみた場合のアイデア数増減の傾向を表 6.4 に示す.

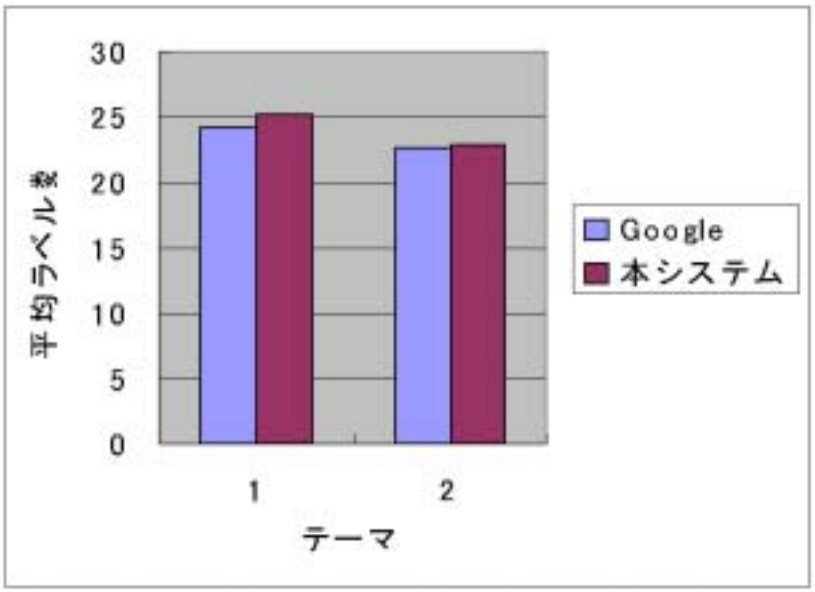

図 6.3: 定量的評価の結果

表 6.4: ラベル枚数で見た場合の傾向

Google 使用時 (テキストデータ中心) のアイデア生成数	本システム (テキストカード+静止画像，図表) 利用によるアイデア生成数の増減
30 枚以上	減 4 人
20 枚以上 30 枚未満	増 1 人，減 6 人
20 枚未満	増 6 人，減 2 人，そのまま 1 人

6.5 考察

図 6.2 を見てみるとテーマによって定性的評価が分かれている．今回の実験では，テーマ 1 では示したデータが画像中心であった．またテーマ 2 においては示したデータが図表中心であった．そのために画像が被験者に直観的に訴えたか否かが評価に差をもたらしたのではないかと思われる．

表 6.4 を見てみると，Google を利用した実験ではアイデア生成数が少なかった被験者が，本システムを利用することによりラベル数が増加する傾向が見られた．これは，テキスト中心での発散的思考が苦手なユーザには，本システムの画像を利用したことにより，発想支援効果が出たのではないかと思われる．しかしことことに関してはさらなる調査が必要である．

第 7 章

結論

7.1 まとめ

本論文では京大型カードを有効に利用できる発想支援ツールの構築について述べた．本研究で得られた成果を以下に示す．

- 京大型カードに基づいた発想支援ツールを構築した
- 評価の結果，文字データだけではなく，画像データを利用した発想支援ツールが有効であるという傾向が見られた

7.2 今後の課題

問い合わせ速度の更なる高速化　本システムは CGI のプログラム言語として Perl を使っている．Perl はインタプリタ言語なためコンパイル言語にくらべ実行速度が遅いという欠点がある．よって本システムの CGI のソースを C 等のコンパイラ言語にすることにより実行速度を向上させることができると思われる．

また本システムは Web サーバとして Apache を使用しているため，mod_perl を使うことも候補として挙げられる．mod_perl とは Perl を Apache のモジュールとして動かす仕組みのことである．一番最初に Perl のプログラムが起動されたときにコンパイルされメモリにキャッシュされる．それ以降はプログラムを起動する時間が省略される．したがって通常 Perl を用いた CGI よりも高速での処理が可能となる．

インタフェース部の改良　問い合わせ結果表示画面に京大型カードのタイトルを表示させるようにして、ユーザにある程度カードの内容が予測できるようにする．

また現在京大型カード表示部分では、スクロールさせないと画像が見られなくなっている．画像データは本文とならぶ重要な項目であるので、画像がすぐに見られるようにレイアウトを変更したい．

またカードを表示した時に，表示された文章が長いと文章を読む作業に時間をとられてしまう．重要文書を抽出したものを出力し，文章のおおよその意味が伝わるようにする必要がある．文書要約器 Posum[14] などを利用して要約された文章を表示させるようにすれば文書を読むための時間が短縮され，思考の流れが妨げられる恐れが軽減すると思われる．

機能拡張　現在は XML ファイルにデータを直接書き込んでいるが，データ管理をデータベース管理システムを行なうようにしたい．また，データベース管理システムを利用した検索機能やデータ入力機能も作りたい．

検索機能は，タイトル，本文，日付やカード作成者を検索キーワードとして検索する機能をつけたい．

現在データ入力は，テキストエディタを用い直接 XML ファイルにデータを入力しているが，フィールドワーク中などノート PS や PDA などモバイル機器からのネットワーク経由での入力を可能にしたい．

謝辞

本研究を行なうにあたって，お世話になった多くの方々にこの場を借りて感謝の気持ちを表したいと思います.

指導教官である國藤進教授には，本研究を進めるにあたっての適切なご指導や助言をいただいただけではなく，さまざまな研究活動のチャンスを与えていただいたことをはじめ，日ごろの研究生活全般に関するご指導をいただき，大変感謝しています.

また，藤波努助教授には，研究にあたって有益なご指導と助言をいただきましたことを感謝いたします.

また，たいへんお忙しい中，長時間にわたる評価実験に精力的に協力してくださいました知識科学研究科の方々に感謝いたします.

國藤研究室の方々には，研究面に限らず，私生活の面においてもたいへんお世話になりました．特に同期生の皆様への感謝の気持ちは絶えません.

また，電子 KJ 法を試用させてくださいました，川喜田研究所に感謝いたします.

ほかに，学会などでさまざまな機会お世話になった方々に感謝いたします.

最後に私事で恐縮でございますが，これまで大学院生活を金銭面・精神面から支えてくださいました両親に感謝の意を表させていただきます.

参考文献

[1] 川喜田二郎: 発想法，中央公論社，(1967).

[2] 梅棹忠夫: 知的生産の技術，岩波書店，(1969).

[3] 國藤進: 発想支援システムの研究開発動向とその課題，人工知能学会誌，Vol. 8，No. 5，pp. 552-559(1993).

[4] 由井薗隆也，宗森純，長澤庸二: カード型データベースを持つ KJ 法一貫支援グループウェアの開発と適用，情報処理学会論文誌，Vol. 39，No. 10，pp. 2914-2926(1998).

[5] 松本裕治，北内啓，山下達雄，平野善隆，松田寛，高岡一馬，浅原正幸: 日本語形態素解析システム『茶筌』 version 2.2.7 使用説明書奈良先端科学技術大学院大学，(2001).

[6] Bray, T., Paoli, J. and Sperberg-McQueen, C. M.: Extensible Markup Language (XML) 1.0 W3C Recommendation(1998).
http://www.w3.org/TR/1998/REC-xml-19980210

[7] Adler, S., et al.: Extensible Stylesheet Language (XSL) 1.0 W3C Recommendation(2001).
http://www.w3.org/TR/xsl/.

[8] 李健，金井貴，國藤進：関連文書によってフィルタリングする連想方式情報検索ツールの開発，情報処理学会研究報告，Vol. 2000，No. 26，pp. 95-100(2000).

[9] 徳永健伸: 情報検索と言語処理，東京大学出版会 (1999).

[10] Witten,I. H., Moffat,A., and Bell, T. C. : Managing Gigabytes: Compressing and Indexing Documents and Images, Morgan Kaufmann(1994).

[11] 金井貴，齊藤主税，國藤進: 情報フィルタリングを用いた対話場におけるナレッジマネジメント，人工知能学会 AI シンポジウム 2000，人工知能学会研究会資料 SIG-J-A003-8，pp. 43-48(2000).

[12] 検索エンジン Google.
http://www.google.com/.

[13] 金子修三: テキストマイニング技法を活用した発想支援システムの構築，北陸先端科学技術大学院大学修士論文 (2001).

[14] 望月源: テキスト簡易要約器 Posum version 1.50.2 マニュアル，JAIST Technical Memorandum, IS-TM-2002-002(2002).

発表論文

1. 若江智秀，小林薫，金丸浩士，藤波努，國藤進：Gush My Spot：知識科学研究科における知識創造支援システム，情報処理学会主催マルチメディア，分散，協調とモバイル (DICOMO 2001) シンポジウム論文集，pp. 151-156(2001).

2. 金丸浩士，若江智秀，小林薫，藤波努，國藤進：フィールドワークで集めたアイディアを有効に利用できる野外発想支援システムの構築，日本創造学会第 23 回研究大会論文集，pp. 71-74(2001).

付 録 A

アンケート調査の結果

表 A.1: 質問 1 の回答

被験者	経験の有無
a1	はい
a2	はい
a3	いいえ
a4	はい
a5	はい
a6	いいえ
a7	はい
a8	はい
a9	はい
a10	はい
b1	はい
b2	はい
b3	はい
b4	はい
b5	はい
b6	はい
b7	はい
b8	はい
b9	はい
b10	はい

表 A.2: 質問 2 の点数と理由 (その 1)

被験者	点数	理由
a1	4	
a2	3	CGI のシステムとしては普通
a3	4	ビジュアルで良い
a4	4	普通
a5	2	表示が少し遅かった
a6	2	キーワードの選択にもよるのだろうが, 結果が似かよったり思いどおりの結果にならないことが多かった
a7	4	類似したものを簡単に見つけられたから.
a8	4	
a9	3	あまりこういったかたちの検索法はしたことがないので どうともいえないです
a10	3	検索方法が 2 パターンあって良かった. ちょっと重い?

表 A.3: 質問 2 の点数と理由 (続き)

被験者	点数	理由
b1	2	
b2	5	わかりやすい
b3	2	慣れていないため．ファイルから選択は使いにくい
b4	4	◎比較的曖昧なキーワードで出力できたこと． URL でも関連付で来たこと．×データ量
b5	2	使いかたがいまいち分からなかったせいかエラーが多かった
b6	2	題目が難しかっただけにいれるキーワードも限定されてたため
b7	5	単語ではなく口語で検索できるのが良い． 類似のものをすぐに検索できるのが良い
b8	3	分かりやすかった
b9	5	
b10	4	軽かった
a 群と b 群の平均	3.35	

表 A.4: 質問 3 の点数と理由 (その 1)

被験者	点数	理由
a1	2	URL だけでは開けてみるまで分からない
a2	4	リストだから
a3	3	ちょっと作業がいる
a4	3	元データに依存すると思うので
a5	3	普通だと思います
a6	4	文章だけではなく，画像も表示されるので イメージ形式しやすいと思う
a7	3	類似度 (?) の高いモノを色分けして表示してあると， より見やすいのでは?
a8	3	普通だと思います
a9	3	上記 (著者註：質問 2) と同じです
a10	4	レイアウトが良かった

表 A.5: 質問 3 の点数と理由 (続き)

被験者	点数	理由
b1	2	手間はかかるものの，出来上がれば手書きより見やすい.
b2	5	XML のファイル名を出されても困ります. タイトルでも表示すればどうでしょう
b3	4	タイトル，本文の並びが良い．タイトルが本文の内容を表しているので内容を読むべきかどうかを決めるのには役に立った. 記録者からの情報源は画像のしたで良いと思う.
b4	4	表になっていたので見やすかった 同じフォーマットなので何度も見るときには便利だった.
b5	3	表示は全てそろえてあるので良かったが，最初の画面が URL だけなので，もう少しヒントがあれば良いかなと
b6	2	文章が長く，用いていた画像も抽象的であったため直感的に分かりにくかった.
b7	4	一覧表示で分かりやすかった.
b8	3	絵がさきに出た方がイメージとして受け取れたと思う.
b9	3	特にない
b10	3	すっきりしてるところが◯
a 群と b 群の平均	3.25	

表 A.6: 質問 4 の点数と理由 (その 1)

被験者	点数	理由
a1	5	だいたいラーメンの情報が得られた
a2	4	
a3	5	文と図がでてくる
a4	5	使えた
a5	4	文書を検索できないと不便だと思います.
a6	2	希望するような結果が得られにくかったため，キーワード入力する方がやりやすい．普段 Yahoo!等を多用しているのでそれ用の入力の方が慣ているせいもある
a7	5	とにかく，見つけやすいことが大事だと思うから
a8	5	キーワード文が多くても検索できるのと必要な情報だけ (論文や雑誌のコメント等) が抜けるので良かったです.
a9	5	これは新たなかんじに思えました
a10	4	明確なキーワードが分からなくても文章でごまかせるから.

表 A.7: 質問 4 の点数と理由 (続き)

被験者	点数	理由
b1	4	抽象的，断片的なキーワードから具体的な文章を導ける点が良い
b2	3	特に他の検索ツールとの違いが分からないです
b3	4	キーワードと関係がある文が出てきたので
b4	3	文書検索をあまり使わなかったので
b5	4	使いかたがわかりやすい
b6	4	出力された結果が多数であったため，様々な考え方が得られた ただ内容が類似しすぎてたものが多かったような
b7	5	口語で検索できるのが良い. 入力されているデータが使えるものが多かった
b8	3	分かりやすかった
b9	5	話題が限定されているせいかいい結果得られた
b10	3	もう少し使いかたがあると思われる.
a 群と b 群の平均	4.1	

表 A.8: 質問 5 の点数と理由 (その 1)

被験者	点数	理由
a1	2	ファイル名だけでは何を検索しているかわからない
a2	4	
a3	4	
a4	5	面白かった
a5	4	類似の文書も参考にしようと思った.
a6	*	(未使用のため評価できず)
a7	5	ラーメンの検索で，味のにているものがほんとうに検索できたから．他のものでも検索してみたい.
a8	5	似たものを検索できると知りたい事項に関して深く調べられるので.
a9	4	
a10	4	ランクの高い類似文書を効果的に使えるので良かった

表 A.9: 質問 5 の点数と理由 (続き)

被験者	点数	理由
b1	4	興味のある情報に近い情報を多く見ることができる
b2	3	どこが類似なのかわからなかったです
b3	3	使い慣れていないためか目的とするデータに たどり着けなそうな気がしてあまり使わなかった. グレーゾーンの情報検索には使えそうな気がします
b4	3	文書検索を行うことがあまりなかったので
b5	*	(未使用のため評価できず)
b6	3	あまり使わなかったのでよくわからない
b7	5	1 ついいのをさがせば他のをさがすのが楽で良い
b8	3	文章が少し長すぎた
b9	2	あまり使わなかったから
b10	4 2	発想は良いと思われる．類似度が意図するものか どうかが今回の使用では不明
a 群と b 群の平均	3.72	

表 A.10: 質問 6 の点数と理由 (その 1)

被験者	点数	理由
a1	4	字数が多すぎるかも
a2	4	
a3	5	文と図でヒントや思い付きが出てきた
a4	4	ヒントになるキーワードがあったりした
a5	1	有効なデータがありませんでした
a6	1	「人気のある味」で検索した上位の結果が本当に人気があるのか判断できなかったため利用しなかった．つまりデータから発想を得るというよりむしろ発想の追認という利用をした.
a7	4	ラベルとなる言葉をいくつか見つけられたから.
a8	5	何となく考えていたものが適切な言葉で 可視化されていたので便利だった
a9	5	
a10	5	何も考えずに文書データをそのまま使えた

表 A.11: 質問 6 の点数と理由 (続き)

被験者	点数	理由
b1	5	文書内のキーワードがアイデアを出すのに役に立った
b2	2	あまり良いデータでなかった気がする.
b3	5	エネルギー系の話はあまりネタがなかったので, システムがなかった場合，半分ぐらいしかラベルが 作成できない気がする
b4	4	何となくわかっているが見つからない言葉をさがすのに 便利だった．同じフォーマットなので限られた時間で 作業するときは便利だった.
b5	4	そこからのキーワードで検索した
b6	4	時間が限定されていたため，長い文章は単語の拾い読み 程度であったが，単語の意味から連想できた.
b7	5	ラベルが最初からまとまっていて役に立った
b8	2	
b9	4	いくつか新しいアイディアが出た
b10	2	読みいってしまうので時間内の発想の数がかえって減る 可能性がある．最初に見た文書に思考がひきずられる.
a 群と b 群の平均	3.75	

表 A.12: 質問 8 の点数と理由 (その 1)

被験者	点数	理由
a1	4	自分の好きなタイプのラーメンになってしまった. “売れる” ラーメンとはちがうと思う.
a2	5	
a3	4	20 分ではまあまあ出たほうだと思う.
a4	4	面白い事例があったので
a5	1	6，7 と同じような意見です
a6	2	意図した結果が出力されず，結局利用せずにアイディアを捻出せざるを得なかった.
a7	4	満足のいくアイデアをいくつか見つけられたけど ツールを利用することで自分の中からアイデアを 出すことができなかった.
a8	4	KJ でできた島の内容を充実させることができたので.
a9	5	
a10	3	アイディアを出す前に文書に意識がいってた.

表 A.13: 質問 8 の点数と理由 (続き)

被験者	点数	理由
b1	4	使わないよりは多くの数が出た
b2	2	データが悪かった?
b3	4	知らなかった情報や忘れていたこと等が システムによって得られたので
b4	4	同じテーマでやっていないのでハッキリしたことが…．画像は役に 立ちました．同じフォーマットなのでやりやすかったです.
b5	4	ラベル数が多くなったから
b6	2	題目に対するイメージと検索があてはまらなかったから
b7	5	類似アイデアがまとめやすかった
b8	2	直接的にはあまり役立たなかった
b9	5	満足した
b10	3	コツをつかむ前に実験終了．コツをつかめば，有効に使えるかも.
a 群と b 群の平均	3.55	

質問 9 の各被験者のコメント

a1　画像が最初に見える位置にあったほうがインパクトがあると思う.

用意してあるデータベースが多分全部ラーメン関係だと思う．ラーメン以外のものからうまいいラーメンのヒントになるものがあると本当の意味で新しいラーメンが生まれると思う.

a2　eKJ だめ．インターフェースが良くない

全てキーワードに関連する結果が出現して感心した

システムの応答時間が早かった

a3　文章もキーワードにできてよかった

図も出てきてよかった

a4　類似文書検索では，自分で書いたのはのぞいたほうがよいと思う.

a5　ラベル作りにはまったく役に立ちませんでした.

テーマが問題かも知れませんが本文内容が味にかんすることが多かったので参考にならなかったのかも知れないです

画像に関しても，ラーメンの盛りつけだけではなく，店の様子も欲しかったです.

a6　日常的に調べ事のため web を利用しているが，意図した通りの語句から目的の文書にたどり着ける例は 50%程度である.

このシステムを利用して感じることは意図通りの結果を出さないことには真価が得られないであろうと思う.

むしろ「ラーメン」「おいしい」と入力したらラーメンの画像が 100 枚位出されると，何らかのアイディアの手助けになるのでは…と思う.

a8　google などで抜いてきた情報には，そのコンテンツ作成者の主観が多くのっているので，無駄な情報が多い．その点 DB 型の支援ツールは必要な情報が客観的に詳細に掲載されているのでよかったです.

a9　20 分という時間が長く感じてしまいました．(最初 10 分ぐらいにアイディアが沢山出て 5 分で調べて行くためです)

a10　URL だけでなく概要も示して欲しかった．

b1　同質の情報が多いと，限られた範囲でした発散しない気がする．

b3　XML ファイルからの検索はカードの中にボタンを作っておいて，ワンクリックでできるようにして欲しい．

b4　同じ難易度のテーマでやってみてもよかったです．
　検索エンジン：データが抱負，フォーマットバラバラ
　システム：データが限定フォーマットが統一

b5　画像は図だけだったような気がしましたが…

b6　検索によるアイディアが浮かんだときは文章よりも文章に含まれる単語のほうが多かった．
　検索へのキーワードは入力スペース的に一語しか入力できず (?)Google みたいに絞りこみ検索できないことが辛かった．

b8　もう少し，イメージ的なほうがよかったと思います．

b9　データの品ぞろえが少な目
　分野が限定されすぎる ↔ 限定された分野では十分な結果が得られた．

b10　短時間で多くの発想をうながすような意図のあるブレーンストーミングでの利用には向かないように思える．発想も固定的になってしまう恐れがある．通常のブレーンストーミング後に使うのが有効か?

付 録 B

ブレーンストーミングにおけるラベル数

表 B.1: a 群被験者のラベル数

被験者	実験に利用したツール	
	google(テーマ 2)	発想支援ツール (テーマ 1)
a1	18	15
a2	19	52
a3	12	12
a4	39	20
a5	27	18
a6	21	19
a7	18	22
a8	15	22
a9	25	27
a10	18	14
合計	212	221

表 B.2: b 群被験者のラベル数

被験者	実験に利用したツール	
	google(テーマ 1)	発想支援ツール (テーマ 2)
b1	31	19
b2	24	6
b3	56	39
b4	16	19
b5	11	24
b6	25	16
b7	36	35
b8	27	17
b9	12	18
b10	26	17
合計	264	210