

Title	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
Author(s)	金子, 修三
Citation	
Issue Date	2001-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/734
Rights	
Description	Supervisor: 國藤 進, 知識科学研究科, 修士

修士論文

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

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

金子 修三

2001年2月

要旨

本稿では、発想しようと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法を活用してユーザにヒントを与えることで発想の支援を行うシステムについて述べる。

目次

1	はじめに	1
1.1	本研究の背景	1
1.1.1	創造的問題解決プロセス	1
1.1.2	発散的思考の分類	3
1.2	本研究の目的	5
1.3	本論文の構成	5
2	テキストマイニング	6
2.1	関連ルールの導出アルゴリズム	7
2.2	関連ルールの可視化とフィルタリング	12
2.3	テキストマイニングの研究	14
3	テキストマイニングによる発想支援	16
3.1	ホームページからキーワードデータベースの構築	16
3.2	ヒントデータベースの構築	17
3.3	関連ルールの提示方法	18
4	システムの概要	19
4.1	システムの構成	19
4.2	システムの実装環境	22
4.3	システムの持つ発想支援機能	23
5	評価実験	25
5.1	実験の目的	25
5.2	実験方法	25
5.3	実験結果	28

5.3.1	定量的評価	28
5.3.2	定性的評価	30
5.4	考察	32
6	終わりに	34
6.1	まとめ	34
6.2	今後の課題	34
	謝辞	36
	参考文献	37
	発表論文	39
	付録	39
A	アンケート調査のコメント	40
B	各ユーザのヒントの生成時間とヒント数の変化	48
C	各ユーザの作成した文章	55

目次

1.1	発想プロセス	3
2.1	Apriori アルゴリズム例	10
2.2	apriori-gen 関数の例	11
2.3	単純ルールの表示例	12
2.4	複合ルールの表示例	13
2.5	日本 IBM の TAKUMI	15
3.1	キーワードデータベースの構築	16
3.2	ヒントデータベースの構築	17
4.1	システム構成図	20
4.2	実験システム	22
4.3	ラベルのキーワード入力	23
4.4	ヒントの出力	24
4.5	ヒントデータベースの中	24
B.1	本システム：被験者 1 のアイデア数の変化	48
B.2	本システム：被験者 2 のアイデア数の変化	48
B.3	本システム：被験者 3 のアイデア数の変化	49
B.4	本システム：被験者 4 のアイデア数の変化	49
B.5	本システム：被験者 5 のアイデア数の変化	49
B.6	本システム：被験者 6 のアイデア数の変化	50
B.7	本システム：被験者 7 のアイデア数の変化	50
B.8	本システム：被験者 8 のアイデア数の変化	50
B.9	本システム：被験者 9 のアイデア数の変化	51
B.10	本システム：被験者 10 のアイデア数の変化	51

B.11	ダミーシステム：被験者 11 のアイデア数の変化	51
B.12	ダミーシステム：被験者 12 のアイデア数の変化	52
B.13	ダミーシステム：被験者 13 のアイデア数の変化	52
B.14	ダミーシステム：被験者 14 のアイデア数の変化	52
B.15	ダミーシステム：被験者 15 のアイデア数の変化	53
B.16	ダミーシステム：被験者 16 のアイデア数の変化	53
B.17	ダミーシステム：被験者 17 のアイデア数の変化	53
B.18	ダミーシステム：被験者 18 のアイデア数の変化	54
B.19	ダミーシステム：被験者 19 のアイデア数の変化	54
B.20	ダミーシステム：被験者 20 のアイデア数の変化	54

表目次

1.1	創造的問題解決のプロセス	2
2.1	スーパーの POS データベース	7
2.2	商品の同時購買	8
4.1	maxsup,maxconf の値とルールの数	21
5.1	被験者の経験	27
5.2	定量的評価	29
5.3	定性的評価：質問 4 と理由	31
5.4	定性的評価：各質問の平均値	32
A.1	質問 5 の評価と理由	41
A.2	質問 6 の評価と理由	42
A.3	質問 7 の評価と理由	43
A.4	質問 8 の評価と理由	44
A.5	質問 9 の評価と理由	45
A.6	本システムの質問 10 の回答	46
A.7	ダミーシステムの質問 10 の回答	47

第 1 章

はじめに

1.1 本研究の背景

1.1.1 創造的問題解決プロセス

創造的問題解決プロセスは、表 1.1[1] に示すように多くの人々によりモデル化されてきた。近年計算機を用いた創造的問題解決プロセスが注目されてきたが、これらの多くは計算機による創造的問題解決支援という観点から分析されていない。國藤は人間の創造性を計算機により支援するという観点から、新たに「発散的思考、収束的思考、アイデアの結晶化、評価・検証」という分類を行った。

発散的思考プロセスでは、そもそも問題はなにかを明らかにすること(問題提起)から出発し、提起された問題に対して、関連情報を虚心坦懐に収集し、現状の分析を行う。つぎに収束的思考では与えられた全情報を構造化するなかで、非本質的情報を捨象し、本質的情報を抽出する。アイデアの結晶化では問題の本質を評価し、問題解決に最も有効と評価される仮説を直感的に評価し採取する。発想はこのような発散的思考と呼ばれるアイデアの生成と収束的思考と呼ばれるアイデアの整理を繰り返し行うことによって進められる。図 1.1 に國藤の提案する発想のプロセスを示す。

國藤の分類では発散的思考および収束的思考が人間の創造的問題解決において最も重要なプロセスであるが、近年発散的思考および収束的思考の計算機によるサポートシステムの研究が活発に行われている。また國藤は計算機による創造性支援として発想支援システムを提案し、「発散的思考、収束的思考、アイデアの結晶化」という一連のプロセスのサポートを目指している。

表 1.1: 創造的問題解決のプロセス

川喜多 二郎	ワラス	パース	ブルーナ	市川 亀久	テボノ	ヴェルト ハイマー	國藤進
問題提起	準備	発想	直感的 思考	アナログ 思考	水平	生産的 思考	発散的 思考
現状把握	あたため						収束的 思考
本質追究							
仮説評価 ・決断	ひらめき						
構想計画	評価・ 検証	演繹	分析的 思想	デジタル 思想	垂直思想	再生的 思想	評価・ 検証
具体策		帰納					
手順の 計画							
実施							
結果の 検証							
総括・ 味わい							

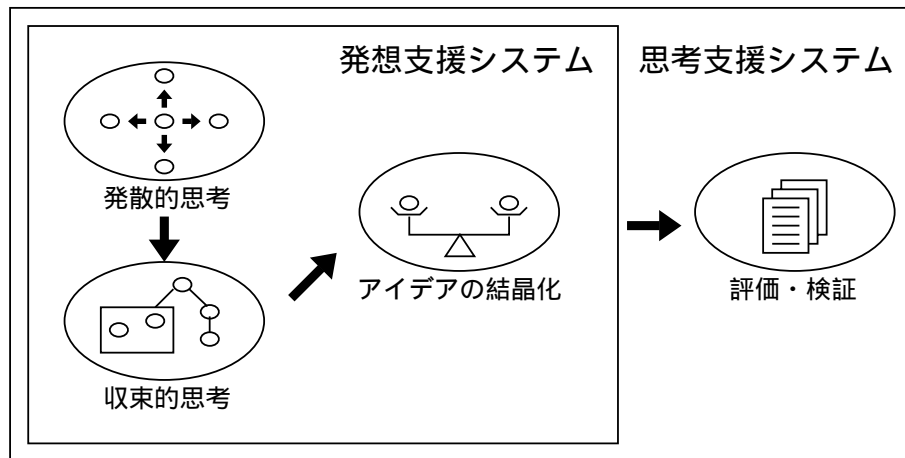


図 1.1: 発想プロセス

1.1.2 発散的思考の分類

計算機による創造性支援研究では「発散的思考, 収束的思考, アイデアの結晶化」という一連のプロセスのサポートが目標であるが, 多くの研究ではすべての思考プロセスをサポートするのではなく, 部分的なサポートを行っているものが多い. 特に発散的思考プロセスに関する研究は数多くなされており, 発散的思考プロセスの分類も行われている. 本節では Young による分類と高橋による分類について述べる.

Young による分類

収束的思考に一貫性やアイデアの実現可能性が求められるのに対し, 発散的思考はとにかく多くのアイデアの断片を生成することが重要視されている.

発散的思考支援ツールとは, 人間の発散的思考を何らかの形で支援する計算機ツールである. しかしこの定義によれば, 発散的思考支援ツールとはきわめて広い範囲を示すことになる. 発散的思考支援ツールは Young[3] によって以下のように分類される.

- ”秘書”レベルの発想支援ツール
- ”枠組み - パラダイム”レベルの発想支援ツール
- ”生成”レベルの発想支援ツール

”秘書”レベルは、基本的に計算機を動的な電子黒板として用いるものである。秘書レベルの発想支援ツールの例としては現在広く使われているワードプロセッサがある。ワードプロセッサはテーマに関連する文章片を思いつくまま入力し、最後に編集するという使い方をされることから発散的思考支援ツールととらえることができる。

”枠組み - パラダイム”レベルは、ユーザが考えをまとめるに当たって適切な枠組みを選んで提供するものである。例としてはワードプロセッサを拡張したアウトラインプロセッサがある。アウトラインプロセッサは文章片の編集機能を強化し、章立て 節立て 内容というトップダウン的なドキュメント作成を支援するツールをいう。

”生成”レベルはユーザが入力した単語を総合したり関連付けることによって、新たなアイデアを生成して提供するものである。このレベルで注目されているのがデータベースからの知識発見であるデータマイニングである。大規模なデータベースを有効に活用するためには、そのデータ間に内在する規則性を見つけ出し、それをルールとしてデータベースに付加してやらなければならない。

本研究ではホームページのテキストデータからデータベースを構築し、新たなタイプの”生成”レベルの発散的思考支援ツールを構築する。

高橋による分類

高橋誠 [16][17] は発散技法を以下の様に 3 つに分けている。

- 自由連想法
- 強制連想法
- 類似連想法

自由連想法はあるテーマについて思いつくまま、次々とアイデアを出す方法でブレインストーミング¹ やブレインライティング法などがこの分類にあたる。

強制連想法はテーマに対して考えるべき方向を示してアイデアを出す方法で属性列挙法やチェックリスト法などがこの分類にあたる。

類似法はテーマと本質的に似たものをヒントとしてアイデアを出す方法で等価変換法やゴートン法などがこの分類にあたる。高橋による分類では本研究は類似連想法にあたる。

¹ブレインストーミングは最も有名な発散的思考技法で批判の厳禁、自由奔放の歓迎、質より量を求める、他人のアイデアへの便乗の歓迎といった 4 つの基本ルールで行われる

1.2 本研究の目的

従来の発想支援システムにおいて、発想は発散的思考と呼ばれるアイデアの生成と収束的思考と呼ばれるアイデアの整理を繰り返し行うことによって進められる。また複数人での発散的思考を行えるケースも多く、この場合他人のアイデアをヒントに新たなアイデアを生成することができる。しかし複数人で発想を行う場合でも参加者の発想がいきづまるとそれ以上アイデアがでなくなるという問題点がある。また「空間の全体」を画面上に表示できないという一覧性の問題、ユーザの入力に対する応答時間をできるだけ早くするという即応性の問題、人間の頭の中にあるメンタルワールドと計算機のディスプレイ上の表示をできるだけ近づけるヒューマンインターフェースの問題がある。

本研究ではこの問題点を克服するために、発散的思考のときに、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法を活用し、ユーザのすでに出しているアイデアと相関ルールのあるキーワードをヒントとして与える発想支援システムを構築する。

1.3 本論文の構成

本論文は6章から構成される。2章では現在盛んに行われているテキストマイニングの関連研究について述べる。3章では実際に本研究で行ったテキストマイニングによる発想支援の方法について述べる。4章では3章で述べた設計をもとに既存の発想支援ツールである IdeaCanvas に実装した実験システムについて述べる。5章では本システムとダミーシステムとを比較した評価実験について述べる。最後に本論文の結論と今後の課題について6章で述べる。

第 2 章

テキストマイニング

計算機ネットワークや WWW の普及によって個人の参照可能なデータの量は飛躍的に増大している。中でも文章形式で蓄積されたデータの増量が目立ち、一説には世の中に蓄積されているデータの 8 割以上が文書形式であるともいわれている。テキストマイニングは、大量かつ多様なデータをさまざまな観点から分析し、役に立つ知識・情報を取り出そうとする技術である。

このような目的のために研究開発された技術としてはデータマイニングがある。データマイニングでは、集められたデータは、データベース・スキーマによってきれいに整理されていることを前提とし、データの特徴的な性質をルール化することを目的としている。

本節では、まずデータマイニングで使われる相関ルールの導出アルゴリズム、相関ルールの可視化、相関ルールのフィルタリングについて説明する。その後現在盛んに行われているテキストマイニングの関連研究について述べる。

2.1 相関ルールの導出アルゴリズム

相関ルール (Association Rule) とは $A \wedge B \Rightarrow C$ (A かつ B ならば C) といったルールのことである。相関ルールには支持度 (support) と確信度 (confidence) という値が付与され、定義は次のようになる。

支持度 相関ルールに含まれるすべての要素がデータベース中に出現する頻度
確信度 データベースにおける相関ルールの正しさを表す

相関ルールの定義

$I = \{i_1, i_2, \dots, i_m\}$ をアイテムの全体の集合とし、 D を処理の集合 (D の要素 T は $T \subseteq D$ を満たす) とする。 D の X を含む要素の $c\%$ が Y も含み、かつ D の $s\%$ が $X \cup Y$ を含む時、ルール $X \Rightarrow Y$ ($X \subseteq I, Y \subseteq I, X \cap Y = \emptyset$) は支持度 $s\%$ 、確信度 $c\%$ の相関ルールと呼ぶ。またルール $X \Rightarrow Y$ の X を前件部、 Y を後件部と呼ぶ。

相関ルールは主にマーケティングにおける情報分析に利用されている。POSを用いることにより商店の膨大な販売情報を収集し、記憶することが可能となった。相関ルールはこれらの販売履歴データベースから販売戦略に役立つ情報を生成するのに適している。

ここで相関ルールを例 [12] を用いて簡単に説明する。表 2.1 は 5 つの商品についての 5 つの処理を示している。それぞれの処理はどの商品が他のどの商品と一緒に購買されているかを示す。このデータから、同時購買された商品のペアの数を示す同時購買表 2.2 を作成することができる。

表 2.1: スーパーの POS データベース

顧客	商品
顧客 1	オレンジジュース, 炭酸飲料
顧客 2	牛乳, オレンジジュース, ガラスクリーナ
顧客 3	オレンジジュース, 洗剤
顧客 4	オレンジジュース, 洗剤, 炭酸飲料
顧客 5	ガラスクリーナ, 炭酸飲料

表 2.2: 商品の同時購買

	オレンジジュース	ガラスクリーナ	牛乳	炭酸飲料	洗剤
オレンジジュース	4	1	1	2	2
ガラスクリーナ	1	2	1	1	0
牛乳	1	1	1	0	0
炭酸飲料	2	1	0	3	1
洗剤	2	0	0	1	2

同時購買表 2.2 では単純なパターンが見て取れる .

- オレンジジュースと炭酸飲料は, 他の 2 つのアイテムよりも一緒に代われやすい
- 洗剤はガラスクリーナや牛乳とは一緒に買われない
- 牛乳は炭酸飲料や洗剤とは一緒に買われない

これらは相関ルールの例であり, 「もし顧客が炭酸飲料を買うと, オレンジジュースも買う」といったルールを示している . 次にこのルールがどのくらい良いルールかを考えてみる .

このデータでは 5 つの処理のうち 2 つで炭酸飲料とオレンジジュースが一緒に出てくる . この 2 つの処理はルールを支持している. これをパーセンテージに表すと支持度は $2/5$ つまり 40% となる.

次に炭酸飲料を含む処理はオレンジジュースも含んでいるので, このルールは確信性も高い! 炭酸飲料を買えばオレンジジュースも買う」のは, 炭酸飲料を含む 3 つの処理のなかで 2 つなので確信度は 67% である . その逆の「オレンジジュースを買えば炭酸飲料も買う」のは処理が 4 つの中で 2 つしかないので, 確信度は 50% である .

この例のように相関ルールの導出とは与えられた値以上の支持度と確信度をもったルールを生成することになる .

ここで与えられた値以上の支持度の要素の組¹ をすべて発見することを考えるてみる . 一般的に, n 個のアイテムの組み合わせの数は n 乗に比例するため, データベースの大きさから考えると高速なアルゴリズムが必須となってくる.

次に相関ルールの導出アルゴリズムの中で代表的な Apriori アルゴリズムについて説明する.

¹このような与えられた値以上の支持度の要素の組をラージアイテム集合という

Apriori アルゴリズム

関連ルールを高速に生成するアルゴリズムとして, Apriori アルゴリズムは広く利用されている . 以下に Apriori アルゴリズムを示す .

1. $L_1 = \{\text{large 1-itemsets}\}$
2. for ($k=2; L_{k-1} \neq \emptyset; k++$) do begin
3. $C_k = \text{apriori-gen}(L_{k-1}); // \text{New candidates}$
4. forall transaction $t \in D$ do begin
5. $C_t = \text{subset}(C_k, t); // \text{Candidates contained in } t$
6. forall candidates $c \in C_t$ do
7. $c.\text{count}++;$
8. end
9. $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$
10. end
11. $\text{Answer} = \bigcup_k L_k;$

Apriori アルゴリズムの最初のパスは 1 項目だけのラージアイテム集合を決定するために単純にデータベースから 1 項目の出現頻度 (support) を数える . その後のパスは 2 段階ある .

はじめ ($k-1$) 番目の中から見つけれられるラージアイテム集合 L_{k-1} は apriori-gen 関数によって候補項目集合 C_k の生成に使われる . 次にデータベースから候補項目集合 C_k の支持度が数えられる . 図 2.1 に Apriori アルゴリズムの例² を示す .

次に apriori-gen 関数について説明する . apriori-gen 関数は ($k-1$) 項目のすべてのラージアイテムの集合 L_{k-1} を引数としてとり, k 項目のすべての集合の候補項目集合 C_k を生成する . 機能は次のようになる .

²この例ではデータベースの処理の数が 4 つとなっている . そして支持度が 30% となっているため support の数が最低 $1.2 (= 4 \times 0.3)$ 以上ないと, ラージアイテム集合 L になれない .

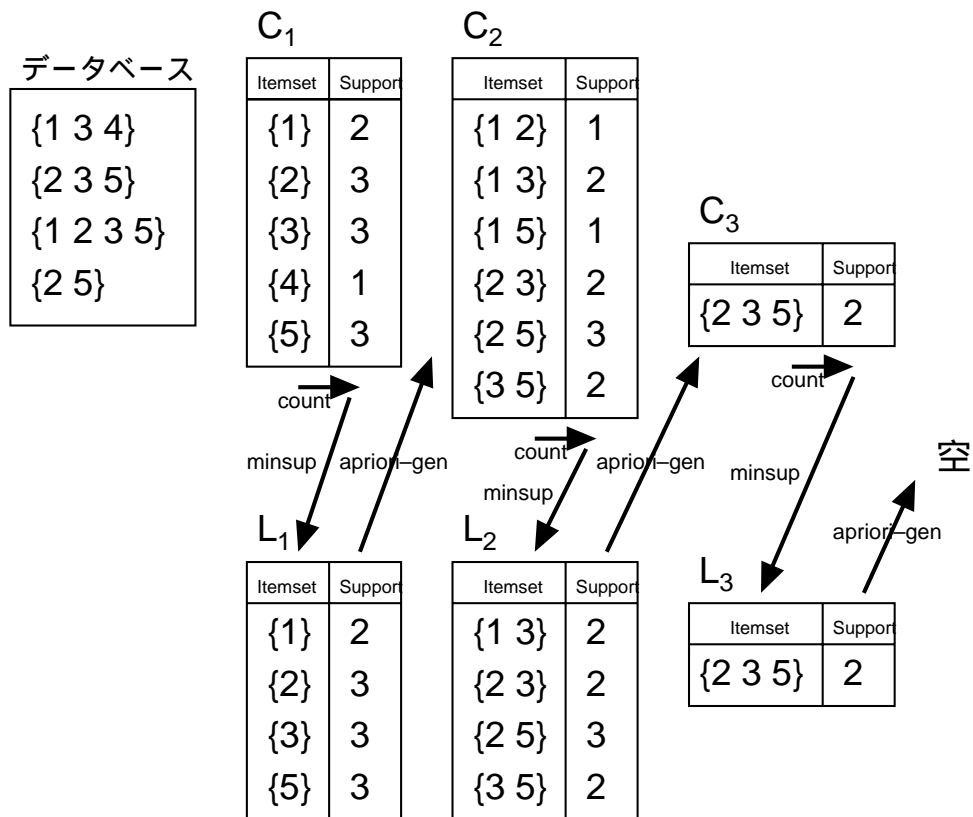


図 2.1: Apriori アルゴリズム例

はじめに join step で L_{k-1} と L_{k-1} を加える .

join step のアルゴリズム

```

insert into  $C_k$ 
select  $p.item_1, p.item_2, \dots, p.item_{k-1}, q.item_{k-2}$ 
from  $L_{k-1}p, L_{k-1}q$ 
where  $p.item_1 = q.item_1, \dots, p.item_{k-2} = q.item_{k-2}, p.item_{k-1} < q.item_{k-1}$ 

```

次に prune step で候補項目集合 C_k の中で (k-1) 項目の部分集合がラージアイテム集合 L_{k-1} の中にないようなすべての項集合 $c \in C_k$ を削除する .

prune step のアルゴリズム

```
forall itemsets  $c \in C_k$  do
  forall (k-1)-subsets  $s$  of  $c$  do
    if ( $s \notin L_{k-1}$ ) then
      delete  $c$  from  $C_k$ 
```

apriori-gen 関数の例

apriori-gen 関数の例を図 2.2 に示す。図 2.2 では join step の後、候補項目集合 C_4 は $\{\{1\ 2\ 3\ 4\}, \{1\ 3\ 4\ 5\}\}$ となる。prune step では $\{1\ 3\ 4\ 5\}$ の部分集合の中で $\{1\ 4\ 5\}$ と $\{3\ 4\ 5\}$ がラージアイテム集合 L_3 中にないので $\{1\ 3\ 4\ 5\}$ は削除される。そして候補項目集合 C_4 は $\{1\ 2\ 3\ 4\}$ だけが残される。

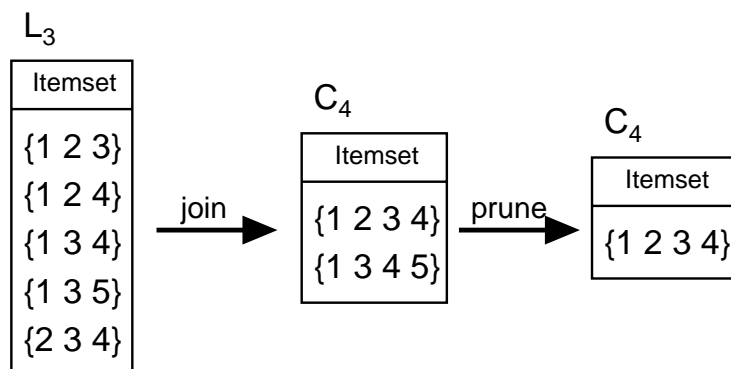


図 2.2: apriori-gen 関数の例

2.2 相関ルールの可視化とフィルタリング

相関ルールが導出されると次の問題として抽出された相関ルールが大量でデータベース全体の関連を一度に把握することが困難になる。そこで導出された相関ルールを有効に活用するために、相関ルールを可視化することが望まれる。さらに次のような問題もある。ルールが確信度を使って選択されるとき「良い」ルール(正しいルール)が必ずしも関連が容易に想像できないルールにはならないことである。例えばルール「妊娠する 女性」が100%の確信度で真実であることを医学的なデータベースで発見するのは容易なことである。しかしこのルールにはあまり驚かない。このような関連が容易に想像できるルールがあると、ルールを見るユーザにとって重要なルールが埋もれてしまい、ルールの理解を妨げとなることがある。そこで相関ルールの導出後にフィルタリングをかけることによってこのような関連が容易に想像のつくルールを削除することを行わなければならない。

日本IBMの福田[5]は相関ルールの可視化とそれにもなうフィルタリングについて考察している。ルールの可視化方法としてはルールを全体として有向グラフに表示する。ルールの前件部、後件部がともに単一のアイテムのみからなる場合は次のようになる。今以下のような3つのルール

- 洗剤 ⇒ オレンジジュース
- オレンジジュース ⇒ 炭酸飲料
- 炭酸飲料 ⇒ オレンジジュース

があるとする。これらを図2.3の様に表示する。



図 2.3: 単純ルールの表示例

前件部または後件部が複数のアイテムからなる場合, 付加的な頂点を使って and を表現する . 例えば

- 洗剤 and 炭酸飲料 \Rightarrow オレンジジュース

というルールがある時, これらを図 2.4 の様に表示する .

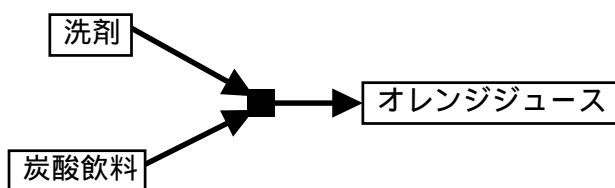


図 2.4: 複合ルールの表示例

ルールの支持度と確信度は, 有効辺の幅 (太さ) と色によってそれぞれ表現する . 即ち, より高い確信度を持つルールを表す有向辺がより太い幅を持つように表示し, より高い支持度を持つルールを表す有効辺がより明るい色を持つように表示する .

また関連ルールのフィルタリングは確信度・支持度を基に以下³ のように行っている .

1. $\text{conf}(A \Rightarrow B) < \text{sup}(R)$ ならば $A \Rightarrow R$ を削除
2. $\text{conf}(A \wedge L \Rightarrow R) \leq \text{conf}(L \Rightarrow R)$ ならば $A \wedge L \Rightarrow R$ を削除
3. $\text{conf}(L \Rightarrow A \wedge R) \geq \text{conf}(L \Rightarrow R)$ ならば $L \Rightarrow R$ を削除
4. $\text{conf}(L \Rightarrow R) \times \text{conf}(L \Rightarrow A) \geq \text{conf}(L \Rightarrow A \wedge R)$ ならば $L \Rightarrow A \wedge R$ を削除

³ここでは支持度を sup , 確信度を conf としている

2.3 テキストマイニングの研究

本節では現在盛んに行われているテキストマイニングの関連研究について述べる。

日本 IBM の TAKUMI

テキストマイニングの機能を実現した例として最も注目されているものに IBM の TAKMI (Text Analysis and Knowledge Mining)[6] がある。TAKMI は、テキストデータの中から概念(キーワードとなる文字列とそのカテゴリー)を抽出し、定型情報と共にさまざまな統計量を計算した上で、その結果をさまざまな観点からユーザに提示する。

TAKUMI は実際に企業のヘルプ・センターのテキスト情報を分析した事例がある。この事例のテキストデータは日本アイ・ビー・エム(株)の PC ヘルプセンターにおいて、顧客からの問い合わせをオペレーターが記録したもので、報告書には機種名、問い合わせ種別(購入相談、要望など)、処理に要した時間などの定型情報のほかに、オペレータが具体的な応答内容を自由な形式の文章で入力したものが含まれている。これらの定型情報とテキストのような非定型な情報から抽出した概念とを組み合わせることでマイニングを行い、その結果を視覚化することが出来る。このツールには傾向分析、トピック抽出、2D マップのような機能もある。

TAKUMI は本研究とテキストマイニングの利用方法が異なっているが大変興味深い研究である。

富士通研究所の ACCENT

富士通研究所は大量のテキスト情報から有用な情報を発見するためのテキストマイニングツール「ACCENT」[7]を開発した。

ACCENT は発想支援システム「HIPS」[8]の分析機能をさらに強化することにより、単語間の連想関係の可視化機能を中心とする連想分析支援ツールとして発展させたものである。ACCENT は、文書群から抽出された単語の間の「連想関係」を、単語の共起性に基づいて計算し、マップ(ネットワーク図)として可視化する。文書を個別に調べてもわからない、文書群全体が持つ特徴・傾向を、この単語の連想マップを通して読み取ることが可能となる。

ACCENT はテキストマイニングを活用した発想支援という点で本研究と共通する部分もあるが、本研究では単語の共起性ではなく相関ルールによって発想の支援を行う。

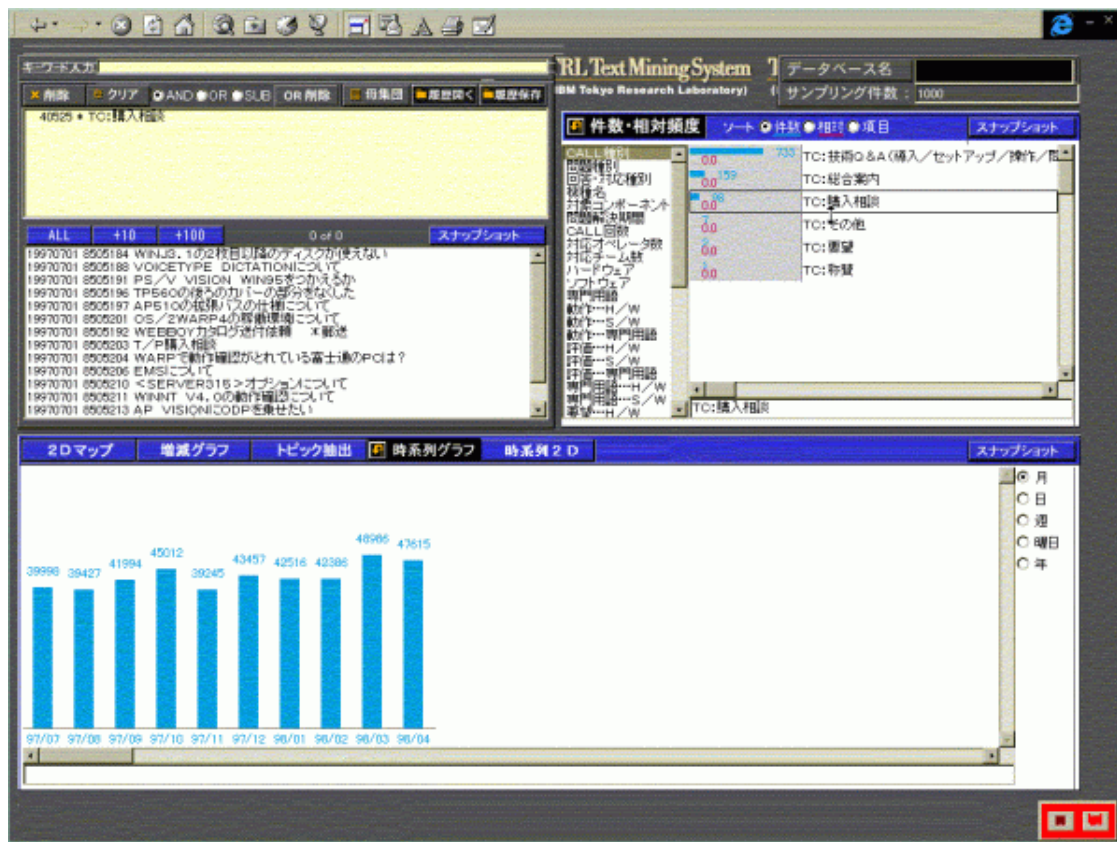


図 2.5: 日本 IBM の TAKUMI

第 3 章

テキストマイニングによる発想支援

前章でテキストマイニングやデータマイニングに関する関連研究を紹介した．本章では実際に本システムで行ったテキストマイニングによる発想支援システムについて説明する．

3.1 ホームページからキーワードデータベースの構築

現在テキスト情報が増大しているが、その中で特に目立つのがホームページである．ホームページはHTMLで書かれたテキストデータであり、これを純粋なテキスト情報に変換しキーワードに切り分けなければならない．本研究ではテキスト情報に変換することをテキストブラウザであるLynx[18]を用い、キーワードへの切り分けは形態素解析ツールである茶筌[19]を用いて行った．品詞に分解されたテーマに関連するテキスト情報は名詞だけを抽出し、キーワードデータベースを構築する．キーワードデータベースを構築する方法を図3.1に示す．

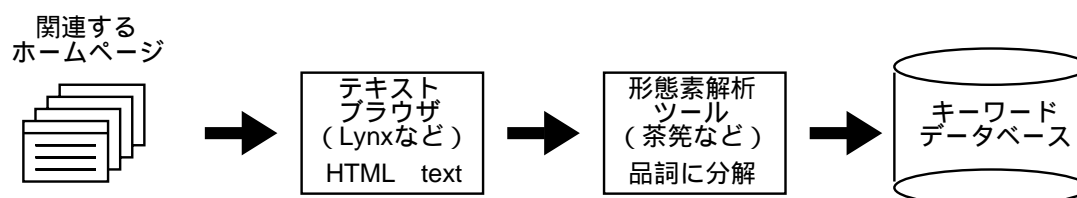


図 3.1: キーワードデータベースの構築

3.2 ヒントデータベースの構築

作成されたキーワードデータベースからヒントデータベースを構築する方法を図 3.2 に示す。ヒントデータベースの構築は, Apriori アルゴリズムを用いて相関ルールを導出し, その中で関連が容易に想像がつくルールをフィルタリングによって削除する。

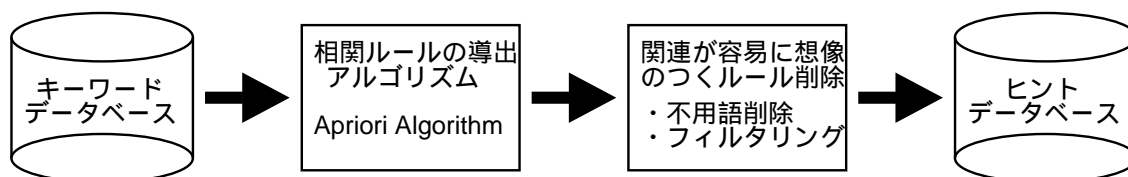


図 3.2: ヒントデータベースの構築

相関ルールのフィルタリング

本研究では以下の 4 つの方法で不要なルールの削除を行った。

- 不要語の削除
- 確信度の近いルールの削除
- 支持度の高いキーワードが後件部にくるルールを削除
- 最大の支持度と確信度の設定

不要語の削除

あきらかにヒントとなり得ない不要語を削除する。削除する不要語は「あなた」「こと」「ため」「よう」「これ」「ただ」「はず」「もの」「ここ」「そこ」「ほう」「ん」「の」「ら」「さ」「つ」「、」などである。このような単語は茶笥を用いたとき, 名詞-非自立語として認識され, キーワードデータベースに残ったものである。

確信度の近いルールの削除

60%の確信度を設定してルール「パン」(確信度 60%)と「チーズ パン」(確信度 62%)を得たとすると2番目のルールは1番目のルールと確信度の違いがほとんどなく、興味深いルールとはいえない。しかしもし「チーズ パン」が確信度 35%の場合はこれは大変興味深いルールかもしれない。なぜならチーズを食べるのならばパンを必要としないということはチーズがパンの代用品であるということをいっているのかもしれないからである。

支持度の高いキーワードが後件部にくるルールの削除

前件部の無いルール(1項目で支持度の高いルール)が後件部にくるルールを削除する。前件部にはできるだけキーワードが多い方がヒントの生成される率が高くなるが、後件部は実際にヒントとして出されるキーワードなので1項目で支持度の高いあたりまえのキーワードはユーザに対する発想の刺激にはなりにくい。

最高の支持度と確信度の設定

相関ルールの導出とは一定の支持度と確信度以上のルールを発見することにあるが、高い支持度と確信度を持つルールは関連が容易に想像のつくルールであることが多い。そこで本システムでは支持度と確信度の上限を明示し、関連の容易に想像のつくルールを削除することを行った。

3.3 相関ルールの提示方法

相関ルールをヒントとして提示するときにはたくさんのルールの中でどのルールを提示するかということが問題になる。本システムでは支持度と確信度の高い順に相関ルールを並べ、そのルールの後件部だけをヒントとして提示する。またヒントの数はユーザが自由に決めることができるようにした。

第 4 章

システムの概要

前章では, 本研究で必要とする発想支援ツールの機能と概要を述べてきた. これらの設計に基づき, インターフェースを近藤[15]の IdeaCanvas をベースとし, 関連するホームページから導出された相関ルールをヒントとして使う発想支援の実験システムを構築した.

本章では作成した実験システムの構成と各機能について述べた後にユーザから見たシステムの利用手順について述べる.

4.1 システムの構成

本システムはクライアント・サーバー方式を採用する. クライアントは図的入力インターフェースを備え, ユーザはその上で発想を行う. 図 4.1 にシステムの構成図を示す. 図 4.1 の上部はテキストマイニング技法を表している.

ホームページのキーワードへの切り分け

本研究では, データベースとして「就職(内定)するには」というテーマでヒントデータベースを構築した. 就職関係のホームページには大きく分けて 3 種類ある.

1. リクルートナビなどの説明会日程などを中心としたホームページ
2. 四季報などの会社データのホームページ
3. 就職活動に対してのこつやアドバイスなどの How to 関係のホームページ

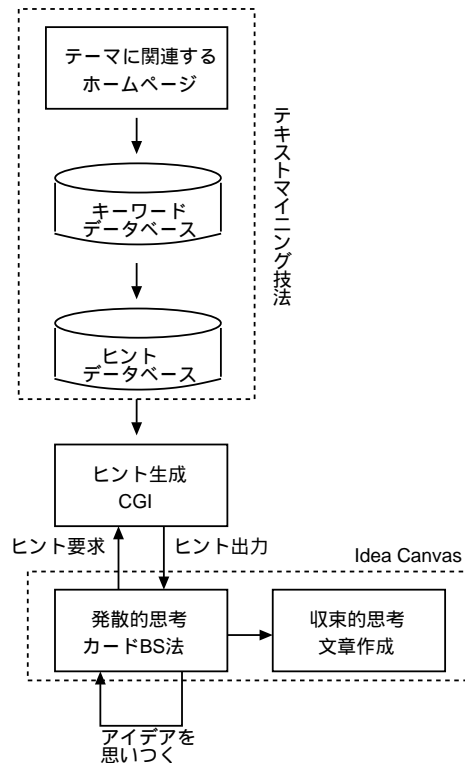


図 4.1: システム構成図

このように単純に goo などの検索を使い、上位いくらかまでのホームページを選択すると今回のテーマに不適切な 1,2 のようなホームページがたくさん選択されてしまう。そこで今回使用したホームページは基本的に就職活動のこつやアドバイスなどを書いてある How to 関係のホームページを選択した。今回使用したホームページ総数は 49 である。

キーワードデータベース構築

キーワードデータベースのスキーマは URL, 文, 単語, 頻度とした。構築されたデータベースの全タプル数は 7920 個で文の数は 1221 個となった。

相関ルールの導出

相関ルールの導出は Apriori アルゴリズムを用いて行った。本研究では一つの文に含ま

れる全単語を一つのタプルとして導出したため, 支持度を 1%, 確信度 5%とたいへん小さく設定した .

ルールのフィルタリング

前章で述べた不要語を削除することによって 11225 個のルールが 7917 個となり, 確信度の高いルールの削除によって 7917 個から 7431 個になった.

さらに試用実験としてフィルタリングに関して支持度と確信度の上限の設定によるルール数の変化について調べてみた . 表 4.1 が maxsup(支持度の上限), maxconf(確信度の上限) の設定によるルール数の変化を表している .

表 4.1: maxsup,maxconf の値とルールの数

maxsup	maxconf	ルールの数	パーセント
なし	なし	7431	100
10	30	7363	99
7	30	7309	98
5	30	7243	97
10	20	7237	97
7	20	7184	96
5	20	7120	96
10	15	6998	94
7	15	6947	94
5	15	6886	92
10	10	6384	85
5	10	6283	84

本研究では相関ルールの上限を支持度 5%, 確信度 10%にした .

4.2 システムの実装環境

これまでに述べたシステム構成に基づいて実験システムを構築した。

開発は Linux¹ 上で行った。ユーザインターフェース部分は JavaApplet で作成された近藤の IdeaCanvas[15] を使い、テキストマイニングを活用したヒント生成部は Perl の CGI によって作成した。ホームページのテキスト化にはテキストブラウザである Lynx を用いた。キーワードの切り出しには日本語形態素解析システムである茶笥を用いた。RDBMS としては PostgreSQL を用いた。



図 4.2: 実験システム

¹Turbo Linux Workstation 日本語版 6.0

4.3 システムの持つ発想支援機能

まず本システムでインターフェースとしている IdeaCanvas の機能を説明する。IdeaCanvas は KJ 法などの実際の展開で模造紙にあたるワークスペースを提供し、ラベルの新規作成、ラベルの選択、ラベルの移動、ラベルの削除・修正などの基本機能がある。本システムではこのような機能にテキストマイニング技法を活用したヒント機能を付加する。

ヒント機能はユーザの発想がいきづまったときにユーザがすでに出しているラベルのキーワードを入力することによって、関連ルールのあるキーワードをヒントとして与える。このことによってユーザが新たなラベルを作成することを助ける。またヒントは必要以上に多く表示されても効果がないためユーザが自由にヒント数を決定できるようにする。

本システムの使用例を図 4.3,4.4 を用いて説明する。まず、ユーザはテキスト入力でテーマに関して思いつくラベルを作成する。思いつくラベルが無くなったとき、すでに出しているラベルのキーワードを図 4.3 の右側にあるヒント生成の部分にヒント数と一緒に入力する。実行ボタンを押すと、入力したキーワードと関連ルールのあるキーワードがヒントデータベースから検索され、図 4.4 のようにヒントが出力される。ユーザは出力されたヒントを参考に再びラベルの作成を行う。この例の時のヒントデータベースの中を図 4.5 に示す。

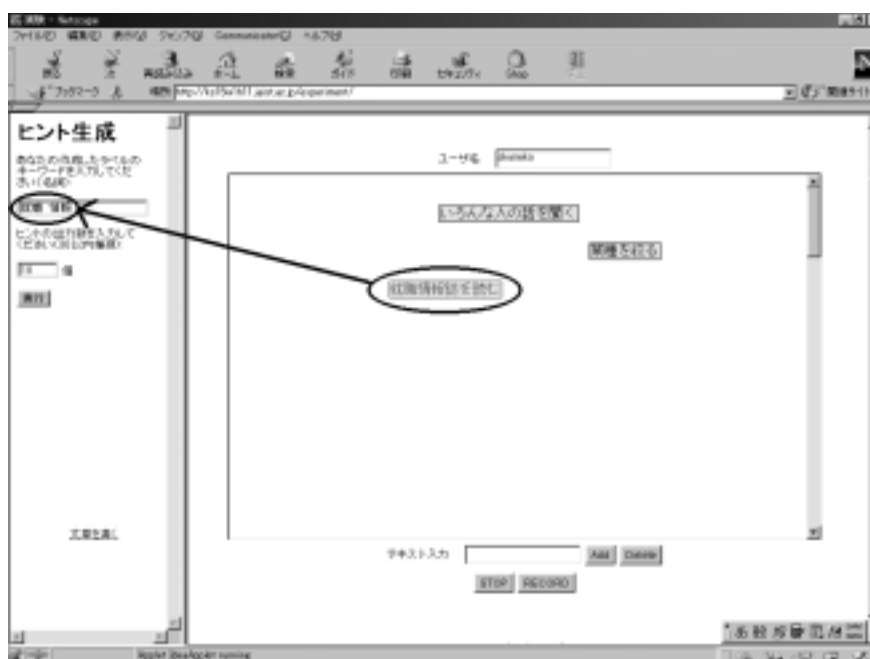


図 4.3: ラベルのキーワード入力

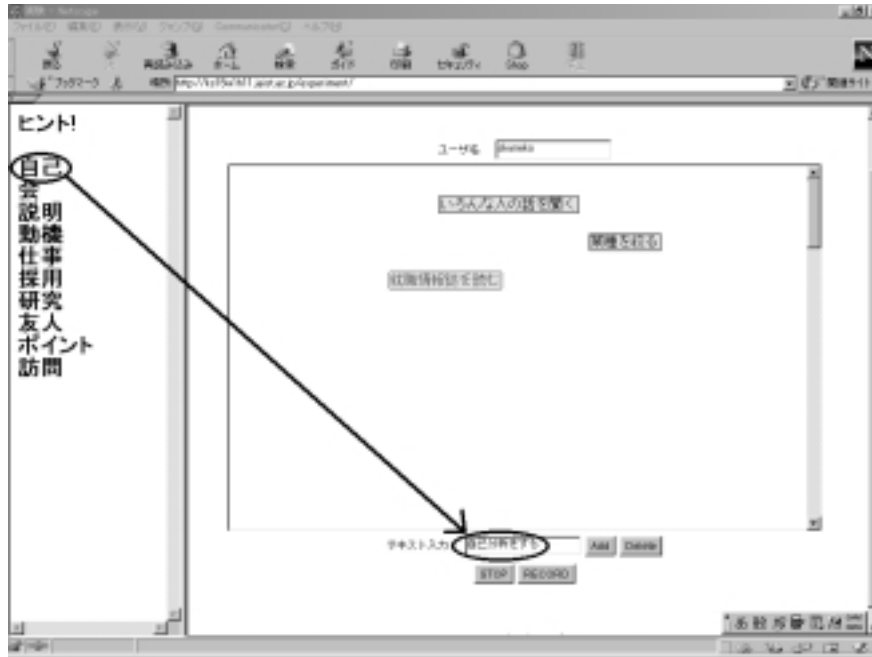


図 4.4: ヒントの出力

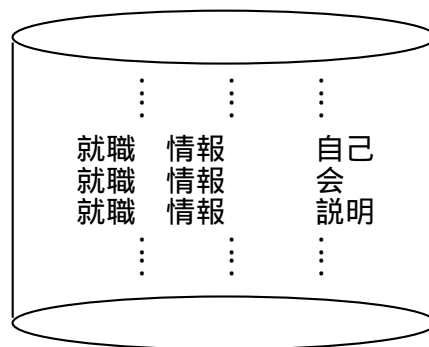


図 4.5: ヒントデータベースの中

第 5 章

評価実験

5.1 実験の目的

本システムは発散的思考で相関ルールを基に作成したヒントを与えることでユーザのアイデアの生成を支援することを目指している。

本実験ではこの効果を調べるために発散的思考を行っているときに本システムのヒント生成を利用してもらいユーザのアイデア生成がどのように変化するかを調べることで本システムの有用性を確認し、さらに今後の機能拡張の方向性を探っていくことを目的としている。

5.2 実験方法

発想支援システムの研究では評価方法として新しい機能があるときとないときの比較によって行うことが多いが、本実験ではヒントがランダムに生成されるダミーシステムを構築し、評価としては偽システムと前章まで述べてきた本システムとの比較によっての有用性を調べる。

実験は、大学院生 20 名を対象にして就職活動について発散的思考を行った。本システムとダミーシステムではユーザの先入観を与えないために同じインターフェースで行った。

実験条件として、10 分の発散的思考をヒント生成を用いながら行った。また発散的思考を終了後、ラベルの空間配置を参考にしながら収束的思考の模擬として簡単に文章にまとめてもらった。評価としては定量的評価としてユーザの入力したアイデアの数、ヒントから生成されたアイデア数、ヒントから生成されたアイデアが文章に使われた数を調べた。定性的評価として被験者にシステム使用後にアンケートに回答してもらった。

各被験者の KJ 法の経験, 就職活動の経験, 今後の就職活動の予定の有無が本実験に影響をおよぼすため, これらの経験について事前の調査を行った。各被験者の経験の有無を表 5.1 に示す。実験は被験者 1~10 に本システム, 被験者 11~20 にダミーシステムで行った。実験を公平に行うために就職活動の経験や今後の就職活動の予定の有無は同じ人数にして行った。

表 5.1: 被験者の経験

	被験者	KJ 法の経験	就職活動の経験	就職活動の予定
本 シ ス テ ム	被験者 1	有り	有り	無し
	被験者 2	有り	有り	有り
	被験者 3	有り	無し	有り
	被験者 4	有り	無し	有り
	被験者 5	有り	有り	有り
	被験者 6	無し	有り	無し
	被験者 7	有り	有り	無し
	被験者 8	有り	有り	無し
	被験者 9	無し	無し	有り
	被験者 10	有り	無し	無し
	1-10 の合計	8 人	6 人	5 人
ダ ミ ー シ ス テ ム	被験者 11	有り	有り	無し
	被験者 12	有り	有り	無し
	被験者 13	有り	無し	有り
	被験者 14	有り	無し	無し
	被験者 15	有り	有り	無し
	被験者 16	有り	有り	無し
	被験者 17	有り	無し	有り
	被験者 18	有り	有り	有り
	被験者 19	有り	有り	有り
	被験者 20	有り	無し	有り
	11-20 の合計	10 人	6 人	5 人

5.3 実験結果

5.3.1 定量的評価

被験者のラベルの生成数, ヒントの生成数, ヒントから生成されたラベルの数, 文章に使われたラベルの数を表 5.2 に示す. また被験者のアイデア数の時間的变化については付録に示す.

表 5.2 を基に下記の採用率を求めた.

本システムのヒントを参考に作られたラベルの採用率

$$\text{採用率} = \frac{43}{400} \times 100 = 10.75\% \quad (5.1)$$

ダミーシステムのヒントを参考に作られたラベルの採用率

$$\text{採用率} = \frac{24}{695} \times 100 = 3.45\% \quad (5.2)$$

本システムのラベルの総数に対するヒントから生成されたラベルの割合

$$\text{割合} = \frac{43}{204} \times 100 = 21.08\% \quad (5.3)$$

ダミーシステムのラベルの総数に対するヒントから生成されたラベルの割合

$$\text{割合} = \frac{24}{191} \times 100 = 12.57\% \quad (5.4)$$

本システムのヒントを参考に作られたラベルが文章へ採用された率

$$\text{採用率} = \frac{16}{43} \times 100 = 37.20\% \quad (5.5)$$

ダミーシステムのヒントを参考に作られたラベルが文章へ採用された率

$$\text{採用率} = \frac{20}{24} \times 100 = 83.33\% \quad (5.6)$$

表 5.2: 定量的評価

	被験者	ラベル 生成数	ヒント 生成数	ヒントから のラベル数	文に使われ たラベル数
本 シ ス テ ム	被験者 1	17	60	8	4
	被験者 2	24	20	3	1
	被験者 3	23	40	5	0
	被験者 4	16	50	4	4
	被験者 5	26	60	9	2
	被験者 6	20	40	2	0
	被験者 7	17	20	5	4
	被験者 8	26	80	5	0
	被験者 9	22	20	2	1
	被験者 10	13	10	0	0
	1-10 の合計	204	400	43	16
ダ ミ ー シ ス テ ム	被験者 11	16	90	2	1
	被験者 12	13	40	2	2
	被験者 13	21	10	0	0
	被験者 14	17	40	2	2
	被験者 15	22	50	5	5
	被験者 16	17	300	2	2
	被験者 17	20	55	4	3
	被験者 18	22	40	2	2
	被験者 19	22	20	4	2
	被験者 20	21	50	1	1
	11-20 の合計	191	695	24	20

5.3.2 定性的評価

アンケートの調査票を以下に示す.

質問 1: あなたはK J法やブレーストーミングをやったことがありますか?

(ある/なし から選択)

質問 2: ヒントをもとに作成したラベルの数はいくつですか?

質問 3: 質問 2 のラベルの中で文章に使われたラベルはいくつですか?

質問 4: ヒントはラベルの作成に役に立ちましたか, 立ちませんでしたか?

(役立つ/役立たない から 5 段階で選択) またその理由もお願いします.

質問 5: ヒントの表示方法は良かったですか, 悪かったですか?

(良い/悪い から 5 段階で選択) またその理由もお願いします.

質問 6: ヒント数を決定できることは良かったですか, 悪かったですか?

(良い/悪い から 5 段階で選択) またその理由もお願いします.

質問 7: アイデアを空間配置できる機能は良かったですか, 悪かったですか?

(良い/悪い から 5 段階で選択) またその理由もお願いします.

質問 8: 全体的なツールの操作感は良かったですか, 悪かったですか?

(良い/悪い から 5 段階で選択) またその理由もお願いします.

質問 9: 本システムはあなたの就職活動について文章をまとめるために役に立ちましたか, 立ちませんでしたか?

(役立つ/役立たない から 5 段階で選択) またその理由もお願いします.

質問 10: その他気づいたことがあれば教えてください.

本評価実験の定性的評価でもっとも重要な質問 4 の 5 段階評価の結果とその理由について表 5.3 に示す. またその他の質問についての結果とその理由は付録に示す. 表 5.4 は各質問の 5 段階評価の平均値を表す.

表 5.3: 定性的評価：質問 4 と理由

	被験者	5 段階	理由
本 シ ス テ ム	被験者 1	4	役に立つものもあったから
	被験者 2	5	ヒントを元にして連想することができたから
	被験者 3	5	ヒントより忘れていた重要項目が想起されたため
	被験者 4	5	自分が気づきそうで気づけなかったところがわかる
	被験者 5	4	発想のヒントになった
	被験者 6	4	想像してなかった単語にきっかけを得ることができた
	被験者 7	5	ヒントによって自分が思いつかなかった発想が浮かんでくること
	被験者 8	3	直感的に役立つものもあるが、間接的に役立つものがあるから
	被験者 9	4	細かなレベルまで考えるきっかけになったが、謎のヒントもあった
	被験者 10	3	
	1-10 の平均	4.2	
ダ ミ シ ス テ ム	被験者 11	2	対応するキーワードが少ない
	被験者 12	2	関連がわからない
	被験者 13	1	ヒントがでてこない。時間がかかる。
	被験者 14	4	
	被験者 15	4	全然関係のないヒントからでも、発想が生まれることがあるから
	被験者 16	2	テーマと入力したワードとの相関が取れたヒントがでない
	被験者 17	4	外からの情報があればひらめきやすい
	被験者 18	2	マッチングしないことが多いから
	被験者 19	4	
	被験者 20	1	あまり関連のないヒントがでてきた
	11-20 の平均	2.6	

表 5.4: 定性的評価：各質問の平均値

質問	本システム	ダミーシステム
質問 4	4.2	2.6
質問 5	3.2	2.6
質問 6	3.6	3.4
質問 7	4.6	4.4
質問 8	4.1	3.3
質問 9	3.8	2.9

5.4 考察

定量的評価として求めた、ヒント数に対するラベルになったアイデア数の採用率をみると本システムがダミーシステムより採用率が3倍近く高くなっているのがわかる。発散的思考はブレインストーミングに代表されるように、質より量が重んじられる点から考えて本システムの有用性を実証しているといえる。また単純にヒントの生成数を見てもこの結果はいえる。

定性的評価ではアンケート調査の質問4で大きな差がでていることがわかる。インターフェースやその他の機能を全く同じにしたことから他の平均点があまり変わらないのに対して特に目立っている。このことも定量的評価と同様に本システムの有用性を実証しているといえる。

またコメントの部分でおもしろい発見がある。本システムを利用した被験者の中に「発想を誘導されているようだった」という意見や「強制連想法のようだった」などがあることである。定量的評価でもいえることであるが相関ルールを用いた本システムはラベルになるまでの採用率は比較的高いが、そのラベルが文章に採用される割合は多くない。逆にダミーシステムはラベルになるまでの採用率は低い文章に採用される割合は比較的高くなっている。これは相関ルールでのヒントが簡単にラベルなることによって被験者が自分でアイデアを考えないため文章に採用されないと思われる。また逆にランダムにヒントを与えた場合、強制連想法のように実際に被験者がじっくり考えてラベルが作成されるので文書に残る割合が高くなると考えられる。

このことは1章の高橋による発想法の分類で本システムが類似連想法のように試用され、ダミーシステムが強制連想法のように試用されたと考えることができる。

次に本研究と関連する研究との比較によって考察する．

関連研究との比較

野口のシステムとの比較

データベースからの知識発見を使った発想支援システムとしては野口のシステムがある．野口のシステムとの違いとしては野口のシステムが収束的思考段階で関連ルールを用いているのに対し、本システムでは発散的思考段階でルールを使っているところである．

これは2つのシステムに使用したインターフェースの違いも関係している．野口のシステムが収束的技法である KJ 法をベースとした富士通研究所の D-Abductor[11] を使用しているのに対して、本システムは発散技法であるブレーンストーミングをベースとした近藤の IdeaCanvas を使用している．KJ 法でもブレーンストーミングのようにラベルの作成は行われるが KJ 法の場合、アイデアの量より質を重視し、根本的にラベルの作られ方が異なる．

またその他では関連ルールのフィルタリングが異なっている．関連ルールの導出はユーザの指定する支持度と確信度以上のルールを見つけることであるが、本システムではそれぞれの上限も作ることによって関連が容易に想像できるルールを削除している．

近藤のシステムとの比較

近藤のシステムは基本的に複数で発想することを前提にしている．そこで同じテーマで他人がまだ発想を行っていないときにヒントが生成できない問題があった．また他人がほとんど同じようなラベルを作成しているときなども有効なヒントが期待できない．

本システムではホームページのテキスト情報を活用しているため一人で発想するときにも有効なヒントを与えることができる．またヒントの数をユーザが決めることができるのでユーザが役に立つヒントだけを活用してラベルを作成できる．

第 6 章

終わりに

6.1 まとめ

本研究では、ユーザの発想がいきづまったときに、発想しようと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法を活用し、ユーザのもうすでに出しているアイデアと相関ルールのあるキーワードをヒントとして与える発想支援システムの設計と実装について述べた。

また、その設計に基づいた発想支援システムの評価実験を行った。評価実験はヒントがランダムに生成されるダミーシステムを構築し、本システムとの比較によって行った。定量的評価では2つのシステムでユーザが作成したアイデア数、ヒントから作成されたアイデア数、ヒントの採用率などを調べた。定性的評価ではシステム使用後にアンケート調査を行った。その結果、本研究で構築した発想支援システムは発散的思考段階でアイデアの生成に有効であることが明らかになった。

6.2 今後の課題

今後の課題としては本システムを発散的思考と収束的思考の一体化技法である統合技法に組み込むことがある。野口はKJ法をベースとした発想支援システムを実装したが、問題点としてちょっとした発想をしようとしている人間がKJ法の図解を作成することに時間を取られ、気軽にシステムを利用しようという気にならない点をあげている。そこでハイブリッジ法などの比較的短時間で簡単に行える統合的発想法に組み込むことが考えられる。ハイブリッジ法は収束的思考で因果分析法やストーリー法といった発想法を用いている。これらの発想法は今回使用した近藤のIdeaCanvasに2章で述べた相関ルールの可

視化をうまく融合させることで、もっと具体的に収束的思考を相関ルールで支援するシステムが構築できると思われる。また IdeaCanvas の既存の機能である他人のヒント機能も活用すればたった一人で発想を行う時だけでなくグループで、しかも非同期での発想支援ツールに拡張することができる。

本システムの課題として、ヒントが生成されないことがある。これはデータベースを構築するときに使用したホームページ数がまだ少ないことや相関ルールを導出するときにある程度キーワードが絞られることが理由と考えられる。またヒント生成のときにユーザが既に出している多くのラベルの中からキーワードを選択することに時間をとられることがある。この問題の解決法としてはある一定の時間ラベルが作成されなければ自動的にヒントを出すようにする等の改良を行う必要がある。また相関ルールのフィルタリングにも課題がある。今回フィルタリングに関してはほとんど検証を行っておらず、本研究が有効なフィルタリングができているか検証が必要と思われる。

謝辞

本研究を行うにあたって、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頂いただけでなく、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御指導を頂き、大変感謝しています。

また、藤波努助教授、金井貴助手には、研究にあたって有益な御指導と助言を頂き、感謝しています。

また、たいへんお忙しい中、長時間に渡る評価実験に精力的に協力していただきました知識科学研究科の方々に、感謝します。

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大変お世話になりました。特に同期生への感謝の気持は絶えません。

ほかに、学会などさまざまな機会にお世話になった方々に感謝します。

最後に私ごとで恐縮ですが、これまで大学院生活を金銭面・精神的面から支えてくれた家族に感謝の意を表させていただきます。

2000年2月13日

金子修三

参考文献

- [1] 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, pp552-559, 1993.
- [2] 折原良平：発散的思考支援ツールの研究開発動向，人工知能学会誌，Vol. 8, No. 5, pp560-567, 1993.
- [3] Young, L.F.：Idea Processing Support, Definitions and Concepts, chapt.8, pp.243-268, in Decision Support and Idea Processing Systems, Wm. C. Brown Publishers, 1998
- [4] Rakesh Agrawal, Ramakrishnan Srikant：Fast Algorithms for Mining Association Rule, Proc. of the 20th Int'l Conference on Very Large Database, pp.487-499, 1994.
- [5] 福田剛志, 森下真一：相関ルールの可視化について，信学技法 DE95-6 pp.41-48, 1995.
- [6] 那須川哲哉, 諸橋正幸, 長野徹：テキストマイニング - 膨大な文書データの自動分析による知識発見 - , 情報処理, Vol.40, No.4, pp358-364, 1999.
- [7] 渡部勇, 三末和男：単語の連想関係によるテキストマイニング，情報処理学会第55回情報学基礎研究会資料, pp.57-64, 1999.
- [8] 渡部勇, 三末和男, 新田清, 杉山公造：ハイブリッド発想支援システム「HIPS」，計測制御学会，第17回システム工学部会研究会「発想支援ツール」資料, pp.77-84, 1999.
- [9] 三末和男, 渡部勇：テキストマイニングのための連想関係の可視化技術，情報処理学会第55回情報学基礎研究会資料, pp65-72, 1999.
- [10] 渡部勇：発想支援システム「Keyword Associator」第二版，計測自動制御学会，第15回システム工学部会研究会資料, pp.9-16, 1994

- [11] 三末和男, 杉山公造 : 図的発想支援システム D-ABDUCTOR の開発について, 情報処理学会論文誌, Vol.35, No.9, pp.1739-1749
- [12] マイケル J.A. ベリー, ゴードン・リフノ著, 江原淳, 佐藤栄作 共訳 : データマイニング手法 営業, マーケティング, カスタマーサポートのための顧客分析, KAIBUNDO.
- [13] 金子修三, 國藤進, 金井貴 : テキストマイニング技法を活用した統合的発想支援システムの提案, 平成 12 年度電気関係学会北陸支部連合大会, pp.444, 2000.
- [14] 野口裕史 : 収束的思考段階の構造を反映して発想の支援を行うシステムの実現, 北陸先端科学技術大学院大学修士論文, 1997.
- [15] 近藤真己 : アイデアの空間配置によるグループ遠隔ブレインストーミングシステムの構築, 北陸先端科学技術大学院大学, 2000.
- [16] 高橋誠 : 創造力事典, モード学園出版局, 1993.
- [17] 高橋誠 : 企画会議の進め方 日本能率協会マネジメントセンター, 1993.
- [18] テキストブラウザ Lynx : <http://lynx.browser.org/>
- [19] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸 : 日本語形態素解析システム「茶筌」 version 2.0 使用説明書, 奈良先端科学技術大学院大学, 1999.

発表論文

1. 金子修三, 國藤進, 金井貴, テキストマイニング技法を活用した統合的発想支援ツールの提案, 平成 12 年度電気関係学会北陸支部連合大会, 北陸先端科学技術大学院大学, 9 月 24,25 日, pp.444,2000.

第 A 章

アンケート調査のコメント

表 A.1: 質問 5 の評価と理由

	被験者	評価	理由
本 シ ス テ ム	被験者 1	3	普通だと思う
	被験者 2	5	左にヒントを見ながら, 右に書くという表示はよかった
	被験者 3	3	
	被験者 4	2	せめて、2 列縦列で並べるくらいの工夫はほしい
	被験者 5	4	じゃまじゃないのでまあまあ
	被験者 6	3	単語の羅列はある意味よいが、もう少し概念じみたものがほしい
	被験者 7	4	単語だったので自分で考えようという気になった
	被験者 8	2	関連度の薄い語句がでるから
	被験者 9	3	何も気になりませんでした
	被験者 10	3	
	1-10 の平均	3.2	
ダ ミ シ ス テ ム	被験者 11	3	
	被験者 12	2	どれがどのくらい関連があるのかわからない
	被験者 13	1	ヒントをあまり使わなかったので評価のしようがない
	被験者 14	2	
	被験者 15	2	「ヒント」という文字が点滅してうざったい
	被験者 16	2	全部ヒントがでない時は, 残りのヒント数も表示してほしい
	被験者 17	3	
	被験者 18	4	大きめのフォントでわかりやすい
	被験者 19	4	
	被験者 20	3	特に見にくいわけでもなく良かったと思う
	11-20 の平均	2.6	

表 A.2: 質問6の評価と理由

	被験者	評価	理由
本 シ ス テ ム	被験者 1	2	あまり意味はないと思う
	被験者 2	4	参画意識が高まった
	被験者 3	4	
	被験者 4	5	良いと思う
	被験者 5	3	いちいち入れるのは面倒, 最初に一回設定すれば十分
	被験者 6	5	選択できるのは都合がよい
	被験者 7	3	数より内容かな?と感じた
	被験者 8	4	関連の薄い語句があっても困るから
	被験者 9	3	ヒント数を変えなかったのでわかりません
	被験者 10	3	
	1-10 の平均	3.6	
ダ ミ シ ス テ ム	被験者 11	4	
	被験者 12	4	いっぱいでてくると見にくいから
	被験者 13	3	
	被験者 14	4	
	被験者 15	5	100 個もでてきても困るし, 少なすぎても役に立たないから
	被験者 16	1	どちらにしる全部見ることになるから
	被験者 17	4	
	被験者 18	2	あらかじめシステムの的に数を決めてもかまわない
	被験者 19	4	
	被験者 20	3	多くですぎてもいやなので自分で決められて良かった
	11-20 の平均	3.4	

表 A.3: 質問7の評価と理由

	被験者	評価	理由
本 シ ス テ ム	被験者 1	4	整理するのに役立つ
	被験者 2	5	文章をまとめやすかった
	被験者 3	4	収束思考をする際似たものをまとめることで考えやすかった
	被験者 4	5	具体的に論点を視覚化できるので良い
	被験者 5	5	視覚的で分かりやすい
	被験者 6	5	グループ分けで考えを整理できて良い
	被験者 7	5	ラベルを移動させながらアイデアを体系化できた
	被験者 8	3	文書を作るには役だった
	被験者 9	5	
	被験者 10	5	
	1-10 の平均	4.6	
ダ ミ シ ス テ ム	被験者 11	4	アイデアの傾向を調べられる
	被験者 12	4	どれとどれが関連するか自分の中で整理しながら考えられる
	被験者 13	4	後で分類わけできるところが良い
	被験者 14	5	
	被験者 15	5	考えを整理できるから
	被験者 16	4	あと、線や四角などがかければ良い
	被験者 17	5	KJ 法もどき (?) が使えるから良かった
	被験者 18	4	考えていることが整理しやすいので好きです
	被験者 19	5	
	被験者 20	4	同じような考えをまとめておけたので文章を書くとき良かった
	1-10 の平均	4.4	

表 A.4: 質問 8 の評価と理由

	被験者	評価	理由
本 シ ス テ ム	被験者 1	3	良いところもあるし悪いところもある
	被験者 2	5	
	被験者 3	5	
	被験者 4	3	これからバージョンアップを望みたい
	被験者 5	3	
	被験者 6	4	直感的でわかりやすい
	被験者 7	4	操作がわかりやすかった
	被験者 8	4	見やすいから
	被験者 9	5	
	被験者 10	5	
	1-10 の平均	4.1	
ダ ミ ー シ ス テ ム	被験者 11	4	
	被験者 12	4	さくさくと動いたから
	被験者 13	3	
	被験者 14	3	
	被験者 15	4	必要最小限のウィンドウ, 項目しかないため混乱しなかった
	被験者 16	3	細かい不満点はあるが, 合格ライン
	被験者 17	1	文章を書くときにラベルが消えてしまった
	被験者 18	4	特に問題ありません
	被験者 19	4	
	被験者 20	3	シンプルな操作感で迷うことはなかった
	1-10 の平均	3.3	

表 A.5: 質問9の評価と理由

	被験者	評価	理由
本 シ ス テ ム	被験者 1	4	一人でじっくり考えるよりは良いと思う
	被験者 2	5	
	被験者 3	3	
	被験者 4	5	非常に役立った
	被験者 5	2	前から思ってたことだから
	被験者 6	4	役に立ったが、時間が少なかった
	被験者 7	4	空間配置によって文章にまとめやすかった
	被験者 8	3	空間配置できる機能は役に立った
	被験者 9	3	もっと時間をかければ役にたったと思う
	被験者 10	5	
	1-10 の平均	3.8	
ダ ミ ー シ ス テ ム	被験者 11	4	
	被験者 12	2	アイデアを出すのに役に立たなかったから
	被験者 13	3	自分の考えをまとめる為には良いがそれ以上にはならない
	被験者 14	3	
	被験者 15	4	たぶん役に立つと思う
	被験者 16	1	ツールを使うより紙で書いた方が良い
	被験者 17	4	
	被験者 18	2	収束的思考をするのが単なる文章化というのはしんどい
	被験者 19	3	
	被験者 20	3	考えをまとめるのには役に立った
	10-20 の平均	2.9	

表 A.6: 本システムの質問10の回答

被験者	その他のコメント
被験者 1	ツールの使い方のコツをつかむのが少し難しい. ヒントが出てこないことが多いのでそのあたりも工夫が必要だと思う.
被験者 2	思ったより, 使いやすく感じた. 自分にとっては切実なテーマだったので良かった.
被験者 3	
被験者 4	もうちょっと, 画面に工夫があれば, やる気ができる人もいるのでは?
被験者 5	色分け, グループ分け, 改装付けがあると便利. ヒントのボキャブラリーがお粗末. 確かにヒントを見て発想するけど, 操作されている様で気味が悪い.
被験者 6	わかりやすく考えをまとめるにはよかった. だが, 発想支援という意味では, 今回の実験ではキーワードででてきた単語にインパクトのあるものがほとんど出てこなかった.
被験者 7	ヒントを提示してもらえるのは良いが, 何としてもそのヒントを使ってアイデアを出そうとしてしまった. 半分強制連想のような感もあった.
被験者 8	ツールの操作法はわかり易いです. ただ若干実験の手順がわかりにくい部分があった.
被験者 9	
被験者 10	

表 A.7: ダミーシステムの質問10の回答

被験者	その他のコメント
被験者 11	一見関係が無い, キーワードに対してもある程度, 内容をカバーする必要ある気がします.
被験者 12	他の人の就職活動をあらかじめ, 空間配置できる機能があれば良い
被験者 13	文章を書いているいくつかのキーワードが新たに思いついた. ヒントははじめからあまり使う気がし無かった. というのは制限があったので頭の中に無数にある漠然としたキーワードまとめることでヒントを使う必要が無かった.
被験者 14	
被験者 15	書き込みをする画面以外は暗い色にした方が落ち着いて作業できると思う. 発想支援, 特に KJ 法などは落ち着いた雰囲気の方がやりやすいと思うから. ヒント数をプルダウン式に並べる方が良い.
被験者 16	
被験者 17	ヒントを単語じゃなくて, 文章でだしてほしい.
被験者 18	ヒントがほしくても該当キーワードがないことが気になりました. むしろ「ヒントが出てきそうなキーワード」を考えなくてはならないぶん, ストレスになっていた. 文章化は, ただ書いてくださいというのではなくて, ある程度は自動生成するとか, もう少し親切にしてくれてもいいかなと思いました.
被験者 19	文章作成時にラベルが見えなくなり, うまくまとめられなかった.
被験者 20	

第 B 章

各ユーザのヒントの生成時間とヒント数の変化

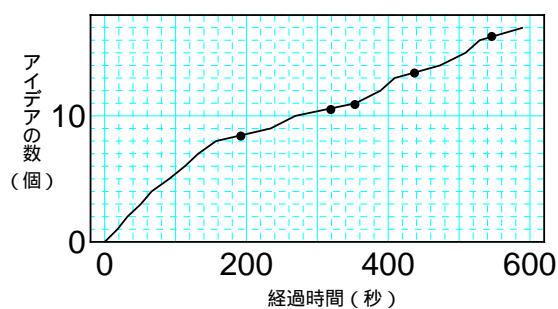


図 B.1: 本システム：被験者 1 のアイデア数の変化

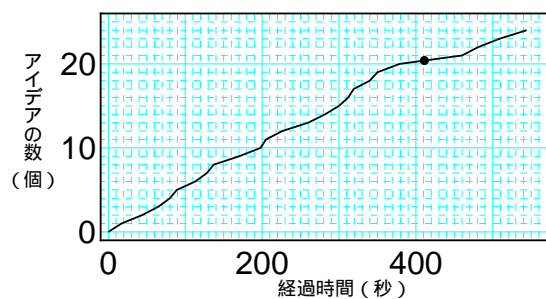


図 B.2: 本システム：被験者 2 のアイデア数の変化

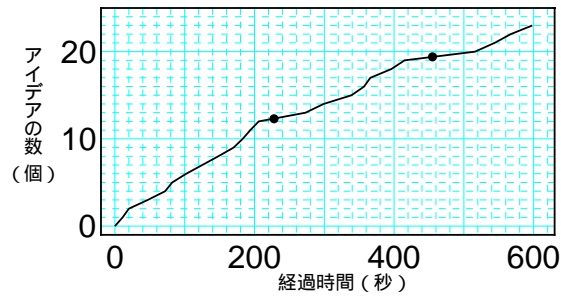


図 B.3: 本システム：被験者3のアイデア数の変化

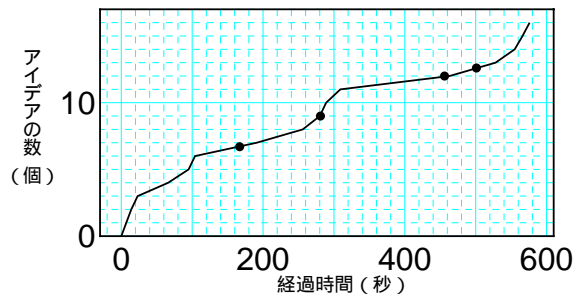


図 B.4: 本システム：被験者4のアイデア数の変化

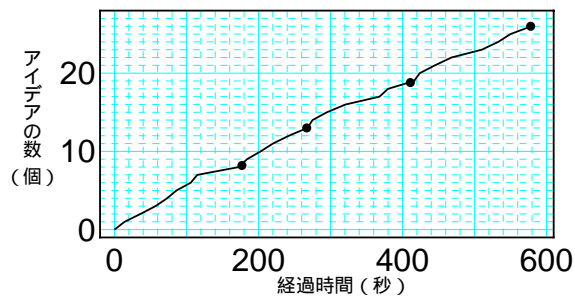


図 B.5: 本システム：被験者5のアイデア数の変化

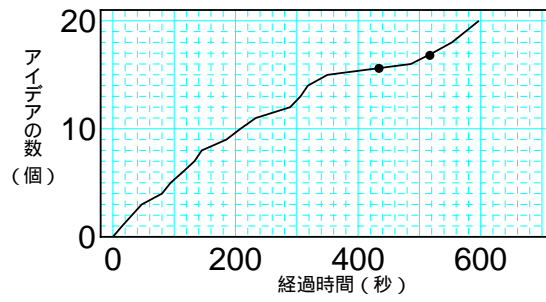


図 B.6: 本システム：被験者6のアイデア数の変化

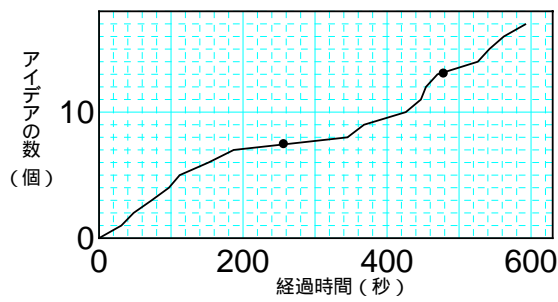


図 B.7: 本システム：被験者7のアイデア数の変化

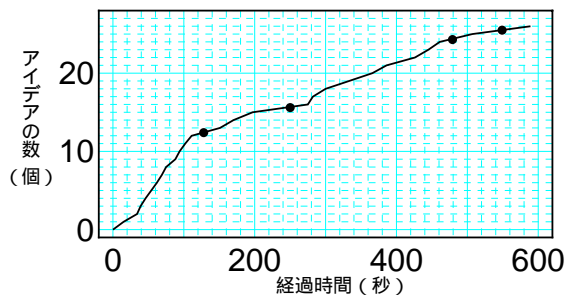


図 B.8: 本システム：被験者8のアイデア数の変化

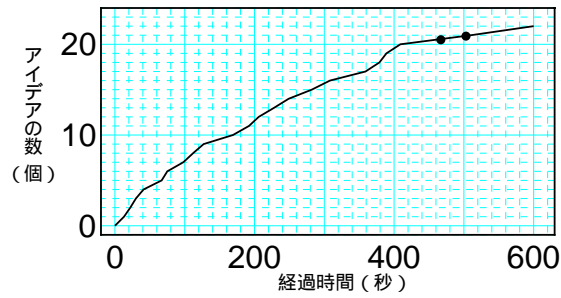


図 B.9: 本システム：被験者 9 のアイデア数の変化

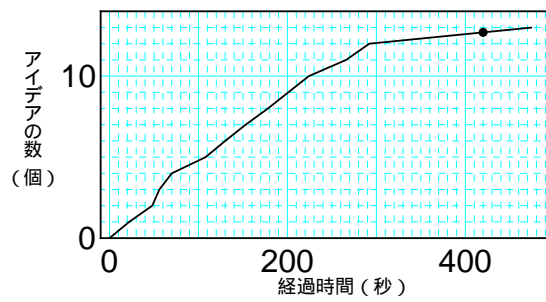


図 B.10: 本システム：被験者 10 のアイデア数の変化

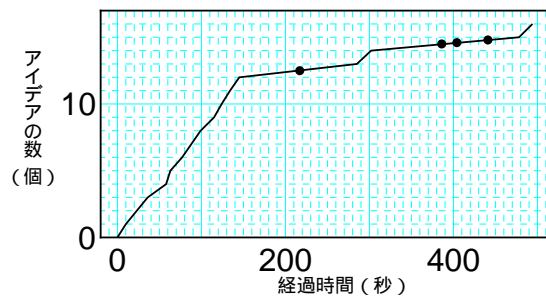


図 B.11: ダミーシステム：被験者 11 のアイデア数の変化

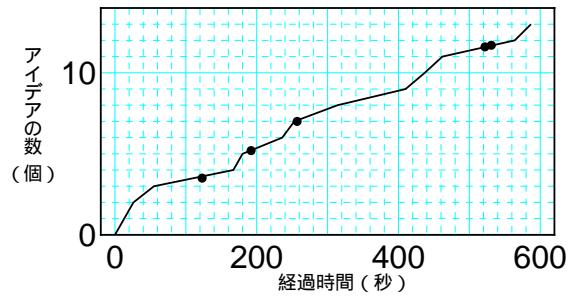


図 B.12: ダミーシステム：被験者 12 のアイデア数の変化

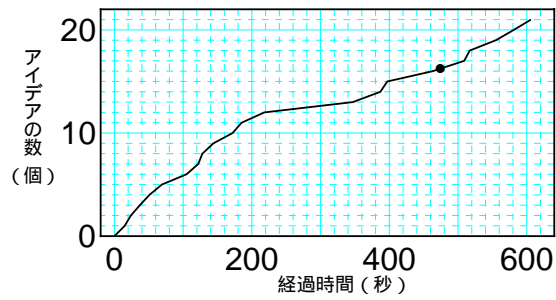


図 B.13: ダミーシステム：被験者 13 のアイデア数の変化

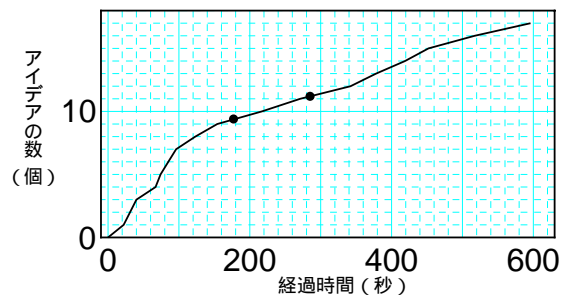


図 B.14: ダミーシステム：被験者 14 のアイデア数の変化

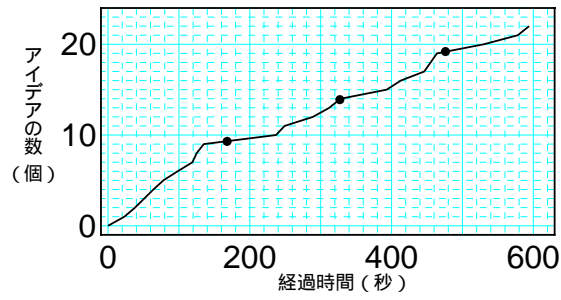


図 B.15: ダミーシステム：被験者 15 のアイデア数の変化

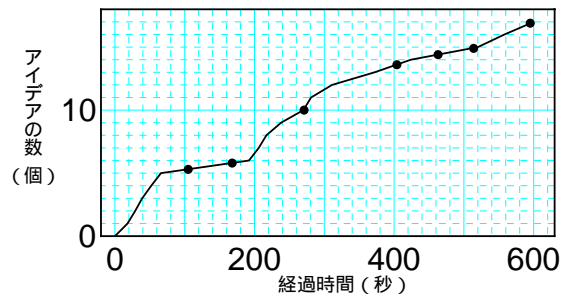


図 B.16: ダミーシステム：被験者 16 のアイデア数の変化

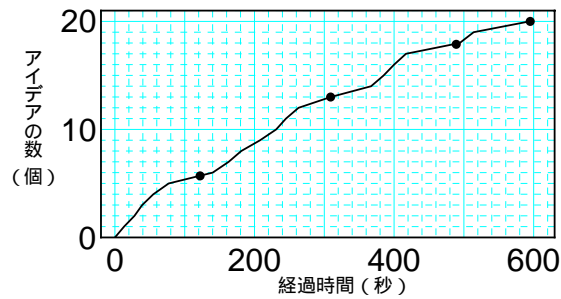


図 B.17: ダミーシステム：被験者 17 のアイデア数の変化

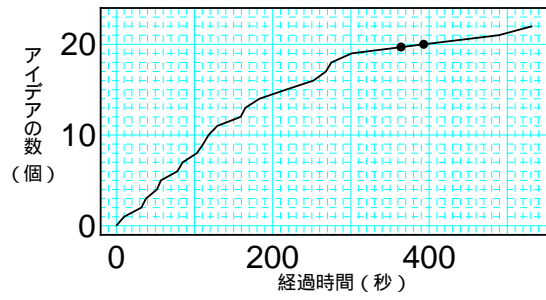


図 B.18: ダミーシステム：被験者 18 のアイデア数の変化

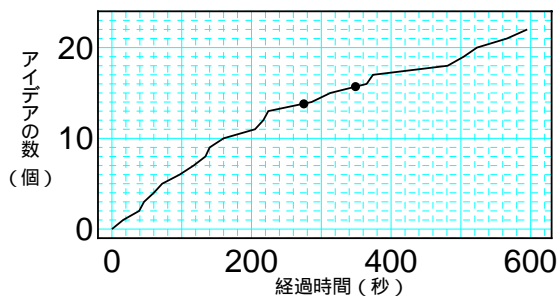


図 B.19: ダミーシステム：被験者 19 のアイデア数の変化

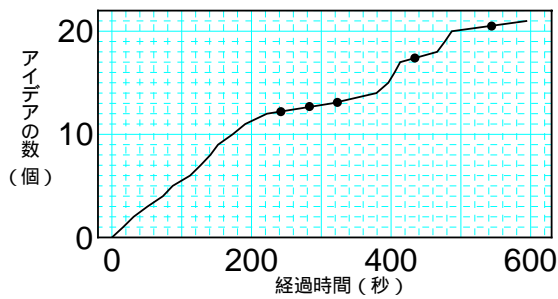


図 B.20: ダミーシステム：被験者 20 のアイデア数の変化

第 C 章

各ユーザの作成した文章

本システム：被験者 1

就職活動はいろいろな人からアドバイスをもらって行動しなければならない。そのときに OB や就職情報誌などで、十分に調べる。実際に活動した時は友達同士で情報交換するのも重要になる。できるだけ早い時期から活動して、ポイントをつかむ。SPI や履歴書などについては早めにつくっておく。何がやりたいのかははっきりさせておいて、志望動機はしっかり考えておく。

本システム：被験者 2

まず、自分が何をやりたいのかをはっきりさせる。そして、企業について調べる。その際に、企業の説明会などに参加する手もある。それと平行して、試験や面接の練習も行なう。自分は、もの忘れがひどいので、スケジュールを付けて就職活動を行なうといいと思う。

本システム：被験者 3

企業と志望者の双方にとって、雇用契約を満足のいくものにするには、志望者は情報収集を怠らず、志望企業について熟知していることが不可欠であり、自己についてよく分析

していることが大切である。

本システム：被験者4

就職活動の重要な点と言えば、エントリーシートを書くことである。その際に、ポイントなるのが、志望動機である。志望動機を活かしてエントリーシートを書く方法は面接の達人などに詳しい書きあがったエントリーシートは友人に見てもらったり、先生に見てもらったりして客観的に見るのが重要である。さて、エントリーシートも重要であるが、その前に自分が志望する企業を決めなくてはならない。その為に、求人広告も見るのも一つの手である。インターネットの時代でもあるから、インターネットで情報を収集するのも悪くない。最近のリクルートなどが、就職専門のポータルサイトを構築しているから一軒の価値ありであろう。OB訪問、会社説明会も有効に利用したい。

最近では、両親の希望もあってか、地元志向の学生が増えているらしい。就職も時代と共に変化していく。できれば、後悔のない選択をしたいものである。

本システム：被験者5

まずやることは業界分析、自己分析、求人情報のチェックをして会社を選ぶ。社会人になって自由な時間が減るのは嫌だし地元も好きなので独立もしくは公務員になって自由に生きたいなあ。

本システム：被験者6

知人に話を聞く。興味のある業界にある会社名を調べる。自分の在籍した学校の先輩の就職先を参考にする。会社のホームページを見て、会社内容、セミナー有無を探る。会社の資料を取り寄せる。Web上での会社評判を検索する。

本システム：被験者7

就職活動を始めるにあたってまず、やるべきことは資料を集めたり、友達とメール交換して情報交換をしたり、インターネットで検索をしたりするなどして情報を集めることである。自らの能力や適性等を考えながら、業界について研究を行なう。業界を把握できて

きたら説明会や会社訪問に行く。自分自身を知り、アピールできるような体制作りとその過程で体力づくりや規則正しい生活を作っていた方が良い。余裕ができたら、マナーや先輩に話を聞く時間を取るなど知識を増やす努力をし、採用試験を積極的に受験すべきだと考える。

本システム：被験者 8

就職活動を始めるにあたり、まずお金、時間が必要です。そしてスーツやネクタイの締め方等多岐にわたります。その後、いざ企業にエントリー、この時には多くの企業でエントリーシートを提出します。そして説明会という順番で進み、さらに面接となります。面接では最近のニュースや集団でのディベート等が難関訓練しておきましょう。また学校推薦も有効です。それには教授とのコミュニケーション、OB や就職課の事務員の人ともコミュニケーションをとっておくと良いです。

本システム：被験者 9

まずは自己分析をしてどんな仕事がしたいかなどをしらなきゃいかんが、どうしたらいいだろうか。どんな会社があるかもわからないといかんのので業種や会社を情報を集めましょう。まずはリクナビ登録。業種や所在地などで検索できるので便利。会社の規模や採用したい学部、勤務地などの条件があわないといけないので要チェック。ホームページはちゃんと見ましょう。(ない時はどうしよう…) だいたい行きたい会社が決まったら PR 文を作ってみる。面接対策だ。大体どんなことが聞かれるかも調べておく必要があるな。面接重視といいつつも適性試験はあるだろうから、その勉強もしておかなきゃいかん。新聞も読みましょう。準備ができたならエントリーです。説明会に申し込み参加です。あとは野となれ山となれ。

本システム：被験者 10

就職活動においては、企業と公務員志望でどちらかを選択する。選択に際しては、絶対安定の公務員を上回る条件をもつ、企業を探す。それには、将来性が不可欠である。次に

企業内の場の雰囲気として、男女の比率も問題にする。

ダミーシステム：被験者 11

自分の就職したい先は、ゲーム会社であるが、おなじような内容を行なえるなら、マスコミ映像関連の会社への就職を考慮しても良い。

しかし、これらの企業は、大部分東京に一極化しており、地方において、これらの職種に就くことは難しいだろう。

また、待遇に就いて考えると、大部分のゲーム会社では、労働組合も組織されていない段階であり、福利厚生が悪い。

だが、現在これらの企業には膨大な応募が集まっており、書類先行にとおるだけでも、難しいものがある。

また、現在すんでいる場所から、東京に就職活動に出る場合、その旅費や、宿泊費用だけで、かなりの額となり、そうとう金銭を消費しそうだ。

よって、ネットなどを利用して、同士との情報交換を行ない、有効な活動を心がけたいと思う。

ダミーシステム：被験者 12

学校推薦をうけるために、担当の先生のご機嫌をとるのも必要。一般で、入るのなら、筆記テストの倍率を勝ち抜く必要があるな。リクルートナビに登録して、まずは企業の会社説明会の日時を管理しよう。もちろん、知人に相談しながら、自己分析をして、面接をがんばらなきゃいかんな。

ダミーシステム：被験者 13

就職活動をするにあたってのまず準備段階として自分がどの職業につきたいかを考える。それにはまず自分のやりたいことを考えたり将来のことについても考えたりする。しかし自分のやりたいことでも自分にとってそれらの職業が適しているかどうかを調べる為に適正テストをすることが望ましい。この様に自分の行きたい会社の業種を絞っていくことから始める。その次に具体的に希望する会社を決めるにあたってすることはまずそれらの業種の会社について調べなければならない。内容、勤務地や給料や休日など、できれ

ばその会社の人に実際に話を聞いてみる事ができれば尚いい。そして希望する会社を決定できれば、実際に活動をはじめるにあたってすることはまず、学校推薦をとれるかどうか調べ、とれれば推薦をとり、できなければ会社のエントリーをしなければならぬ。次にいよいよ面接試験となれば、実際に人事担当の人と合って話をするわけだから身だしなみにも気をつけたほうが良い。もちろん髪などは黒くしておいたほうが印象はいいだろう。また面接で失敗しない為にはそれなりの準備をしておいたほうがいい。就職関係の本を読んだり、先輩の話を聞いたりして面接の練習をすることは重要であると思われる。

ダミーシステム：被験者 14

OB 訪問をしたり会社説明会に行ったりして、企業の感じをつかむ。新聞やホームページなどで資料を集める。先生や親、先輩などに相談する。志望動機をはっきりさせて履歴書を書いてみる。勉強したり資格を取って、自己のレベルをアップさせる。

ダミーシステム：被験者 15

6つの段階があります。

1. 雰囲気をつかむ 2. 情報収集 3. 自己分析 4. 準備 5. 計画 6. 実践

1. 雰囲気をつかむ

就職した、就職活動をした人に話を聞く。これは身近な人に聞くのがいい。また、幸い世間には就職活動を支援する書籍や雑誌があふれているので、それらを読んでもいいと思う。ただ、本の内容にのまれると自分らしさがなくなるので、本の情報からは一歩距離を置く必要がある。

2. 情報収集をする

どんな業界があり、どんな企業があり、社会はどういう仕組みなのかを調べる。行きたい業界をだいたい絞ったら、その分野に関する資料を集めまくって、キーワードなどを整理する。

3. 自己分析をする

情報を集め、頭がパンクしそうになったらそろそろ本気で自分が何をしたいかを考える必要がある。一ヶ月くらい自分探しの旅に出るのも良いと思う。自分の10年後、20年後を思い描いてみるのも良い。

4. 準備

周りも活動を始めだし、そろそろ自分も活動し始めなければやばいと感じだしたら準備

に取りかかる。これは気持ちを切り替えることに重点を置く。一ヶ月間の放浪でうすぎたなく伸びたひげや髪を切り、整え、ビジュアル的に就職活動をやってますという気合を見せる。証明写真を撮り、社会人になった自分を想定して規則正しい生活をする。履歴書を書いてみたりすると気持ちが盛り上がる。

5. 計画

企業のセミナー登録や合同説明会への参加登録をする。遠くで活動する場合は、時間、交通費などの計画を立てる。就職活動用にノートを作り、活動日記をつけていくと良い。その中で自己分析も変わったり、固まったりしてくる。

6. 実践

実際に飛び込んでみる。就職活動中はほかの、活動をしている人に会うので、友達になる。就職活動を通じて自分が得た情報、体験はもったいぶらずに人に教える。説明会、テスト、面接、最終面接をこなすことになるが、なかなか決まらない場合が多い。まわりが決まっても焦らず、あきらめず、「こっちが会社を選んでやる」という姿勢をつらぬく。成功体験、失敗体験をまとめ、自己分析からやり直すことも必要かもしれない。あまりくよくよしないことが大事である。

ダミーシステム：被験者 16

まず、自己分析をしましょう。同時に、身の回りの準備します。

身の回りの準備とは、

- 1 必要な道具を買う
- 2 心の準備をする
- 3 勉強をする
- 4 情報を集める

等です。

情報を集めるとは、友達やOBの人に会ったり、就職関係の本を読んだり、資料を取り寄せたりするのが良いと思います。

そんなカンダで、志望先を見つけたら、エントリーシートとを書いたり、説明会に要ったりしましょう。

すると、試験スタートです。勉強したことと、自分のやりたいこと、はったりをかまして、どんどん先に生きましょう。落ちたとしても、機を弱くしては行けません。

内定が出たら、本島に生きたいところかどうか考えそうでなかったら断りましょう。

さて、内定先はどうですか？

それでは、良い就職活動を...

ダミーシステム：被験者 17

就職活動に挑むにあたって、まず、第一に今、ためこんでる仕事をはやく終わらせる必要がある。それから、リクナビやメーリングリストを活用して外からの情報を積極的にあつめ、同時に自分を内側から見つめなおし、将来の計画をたてる必要がある。あと、留守番電話もかわなくては !! ついでにコードレスにしよう。

ダミーシステム：被験者 18

自分に合った仕事を探すために、まずは自分を見つめなおす。知り合いのをツテ頼り探しながら、情報も集める。情報は耳だけでなく Web や新聞などを見ることはもちろん、自分で歩いて実際に感じて、得る。具体的な行動は、第一印象に気をつかい、自分をしっかりアピールする。やりたい仕事があるが合った会社がない時は、自分で会社を作るのも悪くない。

ダミーシステム：被験者 19

まず自分が何をやりたいのか、どのようなことができるのか、我究する。将来に対する目標が定まったら、その方向性にあった会社を調べ上げる。同時に会社のセミナーに参加したり、新聞を読んだりしながら業界や社会について学び、自分の進むべき選択肢を絞り込む。企業に自分を売り込むために、主テーマのプロポーザルを完成させ、打ち込んでいる研究を話せる様にする。また自分のホームページをまとめ今までやってきたことや、趣味、講義で学んできたことをまとめておく。就職活動にかかる費用を稼いだり、安くアクセスできるような方法を考え、なるべく多くの企業を訪問する。

ダミーシステム：被験者 20

就職活動をするにあたってまずは雑誌を見るなどして情報を集める。やりたいことが決まればどんな企業があるかを探してエントリーする。試験のために対策を立てたりもする。困った時は先生などにも相談し最終的には企業に行って試験を受けることになる。