NAIST-IS-MT9651088

修士論文

WWW 上からの未知単語の概念の抽出に関わる研究

原田 泰吉

1998 年 2 月 13 日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である。

原田 泰吉

指導教官： 西田 豊明 教授
松本 裕治 教授
武田 英明 助教授

WWW 上からの未知単語の概念の抽出に関わる研究*

原田 泰吉

内容梗概

本研究では、WWW 上からの未知単語の概念の抽出に関して論ずる。

WWWには大量で広範な情報が存在する。そしてその情報は常に更新され、辞書に載っていない最新の情報も我々は獲得できる。しかし存在する情報が大量なため、人間の情報処理能力だけで獲得することは困難な状況にある。そこでこの大量の情報の中から概念に関する文だけを抽出する技術が必要である。

自然言語処理の分野では、語彙的な知識の獲得に関する研究が幾つか行われている。そこで本研究では、文中に現れる記述的特徴から意味的関係を獲得するという自然言語処理の分野の技術を応用して概念の抽出を試みる。

これにより手作業で抽出した概念と比較して WWW に存在する概念に関する文を分析し、さらに WWW 上からの未知単語の概念の抽出方法について論ずる。

キーワード

WWW, 概念, 語彙知識獲得

*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 修士論文, NAIST-IS-MT9651088, 1998 年 2 月 13 日.

Study on concept extracting method of unknown word from WWW(World Wide Web)*

Yasuyoshi Harada

Abstract

In this paper , we describe about concept extracting method of unknown word from WWW(World Wide Web).

There are extensive and a large quantity of information in WWW, and these information always renewed. So We can acquire up-to-date information that is not given in any dictionary. But because of a large quantity of information, we are difficult to acquire only the sentence which telling of concept.

In the computational linguistics, some study of acquisition of word sense knowledge have been done.We try to aquire the concept by using the extracting method of semantic information.

we can analyze the sentence of concept which exsists in WWW by comparing the sentence of concept which extracted by the extracting method of semantic information with that which extracted by man.Further more,we describe extracting method of unknown word from WWW.

Keywords:

WWW,concept,word-sense acquition,

---

*Master's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9651088, February 13, 1998.

目 次

図 目 次

表 目 次

第 1 章

序論

1.1 はじめに

本研究では WWW(World Wide Web) の情報量を利用した未知単語の概念の抽出方法についての基礎技術の研究を行なう。

コンピュータの登場以前、人間は自分の知らない事について知ろうとする時、それについての情報は辞書やそれに関して書かれた書物、あるいは他人に聞くことでしか獲得することができなかった。それは人間の脳と紙という媒体でしか情報を記録しておけなかったからである。

しかしコンピュータの登場によりこの状況は変わってしまった。コンピュータは情報を磁気という形で保存することを可能にし、新しい記憶媒体になったのである。さらに近年、情報技術の進歩により通信速度の向上、安価で高性能なコンピュータの生産が可能となった。これにより多くの人間がコンピュータを利用するようになった。さらに電話会社の通話料金の値下げにより、人々は E-mail といったインターネットに参加する機会が増大した。人々の参加にともない、インターネットでの情報は爆発的に増えた。

これにより我々はインターネットという媒体から必要な情報を入手することが可能になった。しかしそこに存在する情報は図書館のようにインデックス別に情報が置かれているわけではない。また情報の量があまりにも多く一人の人間が処理することは困難な状況にある。この困難を克服するために、我々は情報処理技術の開発で処理能力を高める必要がある。そのために本研究は WWW 上に存在する情報 (テキスト) を分析し、情報獲得の方法を提案を行なう。よって本研究は情報処理技術の開発を行なうための基礎となるものである。

1.2 本論文の構成

本論文は序論も含め全5章から構成されている。2章以降の概要はは次のようになっている。

第2章では本研究で取り扱う未知単語、概念の定義付けを行なう。さらに本研究の背景である従来の未知単語の概念の獲得に関する問題点を指摘する。最後に概念抽出技術に関して述べる。

第3章では本研究で行なった機能表現による概念抽出方法を述べる。

第4章では第3章で提案した方法にしたがって実際に未知単語の抽出を行なう。そして各単語について得られた結果を示す。

第5章では得られた結果から、抽出方法についての考察を行なう。またさらに踏み込んだ抽出方法を提案する。

1.3 本研究の目的と意義

本研究は人間の知識獲得行動を支援する概念抽出技術の研究を目的としている。具体的には WWW 上に存在する情報からユーザが知りたい未知単語の概念の意味を含む文を抽出する技術に関して提案、考察をすることである。ここで概念の意味を含む文とは概念の説明として解釈可能な文のことである。

具体的には概念の意味を含む文の抽出を機能表現を用いた方法を採用し、それに基づいて WWW 上から抽出を行ない、WWW 上に機能表現が存在するのかを調べる。またどのような機能表現が妥当なのか分析し、概念の意味を含む文の抽出方法を提案する。

これにより辞書以外の情報源から未知単語の概念を獲得することが可能になる。

第 2 章

概念抽出技術

2.1 概要

本章では、WWW からの未知単語の概念の抽出技術を提案をする。

そのために本章では最初に本研究における未知単語、および概念の定義を行なう。さらに従来の未知単語の概念を得るために行なわれる作業についての非効率さを指摘し、本研究の必要性を明らかにする。最後に本研究の目的を達成させるための技術について述べる。

2.2 未知単語・概念

本節では未知単語、概念とは如何なるものなのかを定義する。これらを定義することにより本研究がどのような単語を対象として、どのようなものを抽出するのかを明らかにする。

未知単語

未知単語は一人の人間が知らない単語を未知単語とする。例えば「本」という単語について多くの人間が知っていたとしても、その人が知らなければ「本」は未知単語になる。つまり本研究で述べている未知単語とは、社会で十分に認識されていない単語のことではなく、利用者が知らない単語のことをいう。

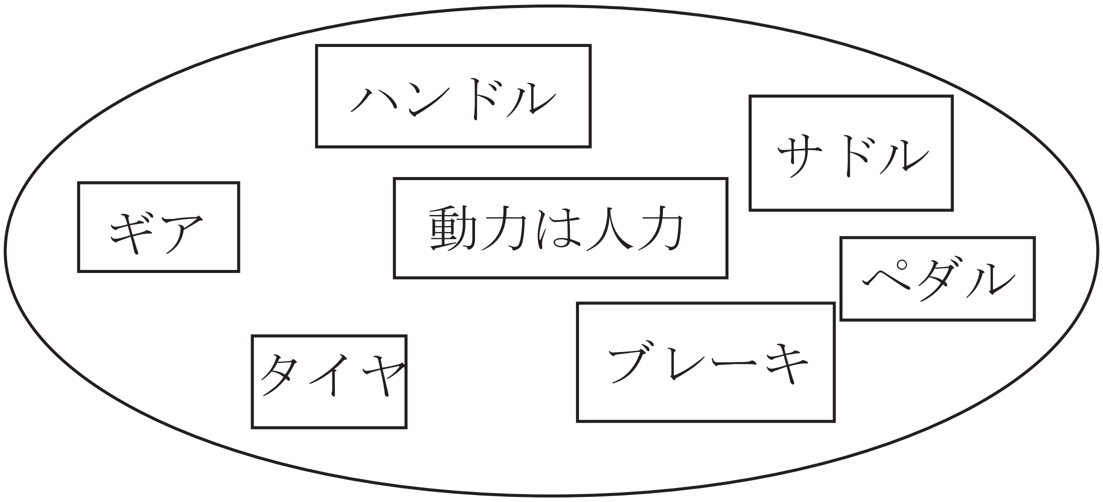

図 2.1　自転車の概念の例

概念

概念とは「何らかの抽象化によって、内界・外界の個別の対象を類別するための知的構成物（体）」であるとされる [1]。つまりある対象とそれ以外の対象とを識別する知的構成物のことを概念という。さらに [JIS Z 8301][2] によると、知的構成物である概念を構成するものは個々の性質、つまり特徴のことであると記されている。概念の例として、自転車の場合を図 2.1に表す。このように自転車に関する特徴を列挙することにより、自転車が何であるのかを知ることができ、オートバイといった他の概念と区別することができる。したがって、ある対象に関する特徴を知ることが概念を知ることになるのである。

2.3　従来の未知単語の概念獲得

概念を獲得するためには、概念を表す情報源が必要である。概念を表す情報源として辞書や他人の知識がある。辞書には単語の意味・情報に関する多くの情報が含まれている。しかし辞書は一度出版されるとなかなか改訂されず、新しい単語に関する概念の獲得は難しい。また他人の知識はその人の知識の量やどの分野に詳しいかに左右されてしまう。

WWW は常に情報が更新され、新しい単語に関する情報も得られるため概念

を獲得する情報源として有効である。しかし WWW から概念を獲得する場合、あまりにも情報が大規模化し、個人の管理能力で情報を収集、理解することは困難である。この問題を解決するために、大量の情報の中から必要なものを取り出す情報検索が一つの分野として確立している [3]。しかし検索できるデータが増えるにつれて検索結果は情報過多になる傾向があり、検索結果を人間が処理するのが困難な状況に陥っている。

またこれらの方法で未知単語の概念を獲得しようとすることは、調べる単語数が少ない場合には調べる時間はそれほど気にならない。しかしオントロジーを作成するような場合、このような方法では非効率である。ここでオントロジーとは、「人工知能システムを構築する時に用いる基本概念(語彙)の体系」[4] である。図 2.2 にオントロジーの例を示す。この例では、implement という概念の下位に fastener や container が存在することを表している。つまり fastener などは implement に包含されていることを表している。このようにオントロジーは概念体系を表しているから、人間がオントロジーを作る時には概念を知っている必要がある。しかし人間一人が全ての対象の概念を知り得ていることは不可能である。また一つの対象、例えば図 2.2のオントロジーを書く場合にも、fastner と container がどう違うのかを知っている必要があるし、implement とはどういうものかなども知っている必要がある。

したがってその分野の専門家以外は、辞書などから多くの概念を獲得してこなければならず、かなりの時間を概念獲得のために必要としてしまう。

2.4 概念獲得技術

本節では概念を獲得するために行なわれてきた従来の研究について述べる。自然言語処理の分野では意味辞書の構築を目的として幾つかの研究が行なわれてきている。意味辞書とは構文解析などの解析技術を実用上有効なものとするために、個々の単語について様々なタイプの意味的あるいは語彙的な知識が記述されたもの [5] である。単語の意味的あるいは語彙的な知識としては、主として次のようなものが挙げられる。

- 概念の集合および概念と表層語との関係

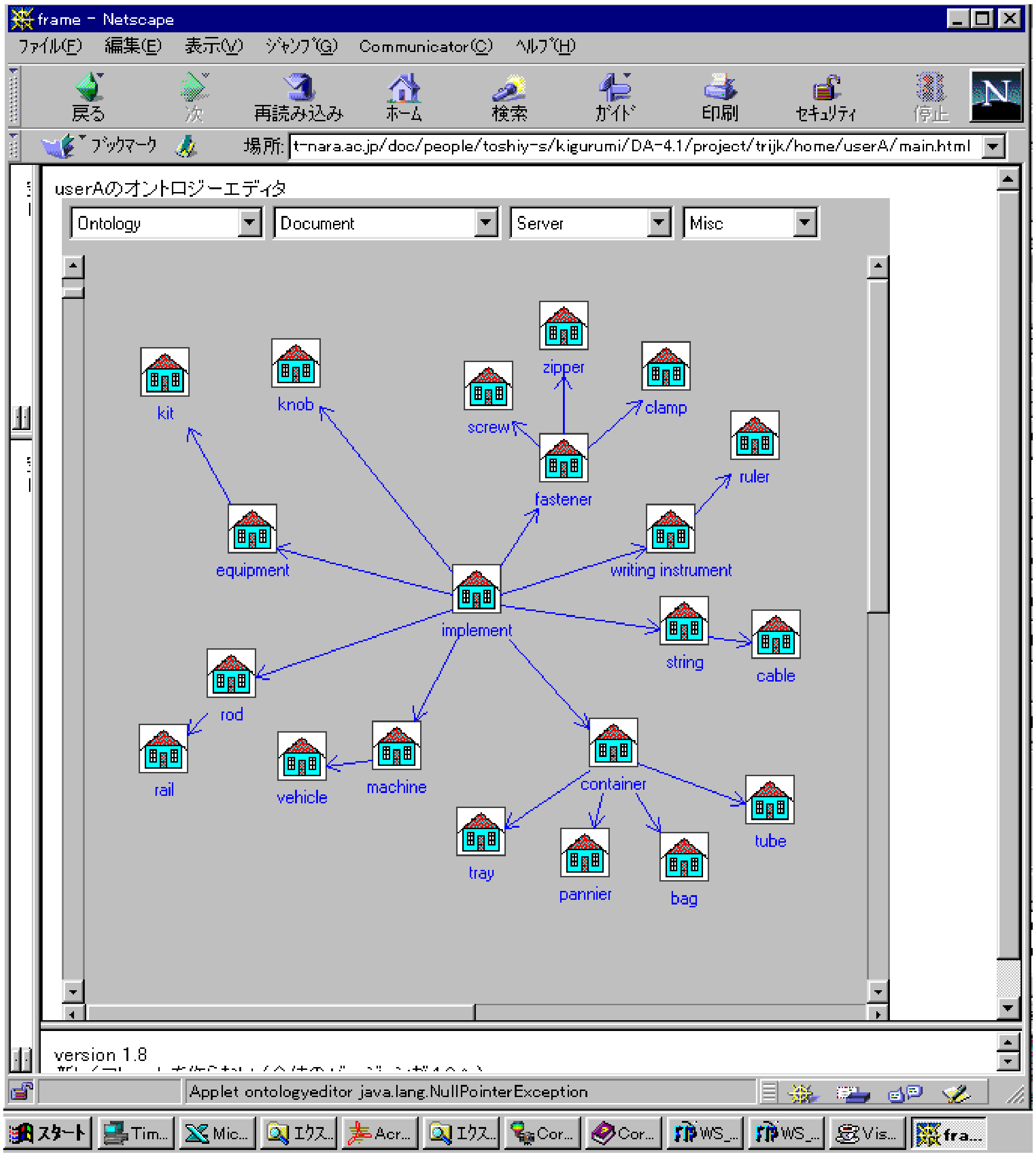

図 2.2 オントロジーの表示例

- 概念間の階層関係あるいはシソーラス
- 名詞概念の持つ属性概念や動詞概念の持つ格要素などの、概念間の非階層的関係

このような意味辞書を、人手で構築するアプローチと、計算機を利用してできるだけ自動的に構築するアプローチがある。しかし人手によるアプローチでは作業の量が膨大であり意味辞書を拡張することも容易ではない。これに対して計算機によるアプローチは人間の負担が軽減され、拡張も容易である。

2.4.1 国語辞典からの階層関係の自動抽出

鶴丸ら [6][7][8] は、国語辞典ではスペースの制限で語義記述を簡潔にするために辞書特有の方法が取られていることに着目した。語義文の中で定義語 (単語の意味を記述 (定義) する場合の中心となる語。例えば {「 鎌 」...草を刈る農機具 } とあった場合農機具が定義語である。) の出現位置や個数に着目した場合の記述的な特徴を次のように挙げている。

(1) 語義文の末尾に定義語が現れる場合が多い。

(2) 語義文の末尾が定義語でなく、特殊な表現が現れている場合がある。これを機能表現と呼ぶ。

(3) 定義語が修飾部を持たない場合がある。

(4) 複数の定義語が ' や'、' と'、'・'、' または' などで結合されている場合がある。

(5) 定義語の直後に例示助詞の' など' が接続している場合がある。

そしてこれらの特徴から近似的に 2 つのタイプに分類している。

タイプ I： 語義文の末尾が定義語である。

例 1 ：「 スピン 」旋回。

例 2 ：「 親王 」天皇の男の子または孫。

タイプ II：語義文の末尾が機能表現である。

例１：「流体」気体と液体との総称。

例２：「海人草」マクリの薬品としての称。

鶴丸らはタイプ I の場合、語義文が見出し語の表す概念と同義な概念を表していて、タイプ II の場合、機能表現が見出し語と語義文との間の意味関係を規定していると仮定して階層関係の抽出を行なった。

ここでタイプ II の場合を詳しくみてみると、機能表現は一般にそれが表す意味関係と密接に関連する語（機能語）を末尾に含んでいる。つまり機能表現は見出し語と語義文との間の意味関係を規定している。また機能表現はそれの表す意味情報に基づいて 2 種類に分類される。

(1) 見出し語と語義文との間の階層関係を積極的に表しているもの（関係情報を持つ）

(2) 見出し語の語用に関する情報を表しているもの（語用情報を持つ）

彼らは関係情報のパターンを見つけ、それらから階層関係を求めている。彼らが求めた機能語を表 2.1に示す。個々で各機能語に着目すると、概念を抽出する機能語として使える語が存在している。例として「一種」という機能語をみてみる。今、「キルトは一種のスカートである」という文があったとする。この文はキルトについての特徴を表していて概念の意味を含む文として有用である。

また彼らは「生産物」の概念に関する見方について図 2.2のように提案している。これは例えば「紙は主に植物性の繊維を材料にしている。」という紙の概念の意味を含む文があった場合、「材料」という表現からこの文は製造に関する見方に制限されるのである。

本研究ではこの「一種」や「材料」という機能表現を用いて WWW からの未知単語の概念の抽出を試みる。

表 2.1 機能語（FW）, 見出し語と語義文との間の意味関係（ρ_{FE}）, および語義文例

機能語 (FW)	ρ_{FE}	備考（語義文例）
略	$\equiv$	[原爆]「原子爆弾」の~
異名	$\equiv$	[エビムスメ] コンブの~
古名	$\equiv$	[あやめ] ショウブの~
改まった言い方	$\equiv$	[家父]「自分の父」の~
一種	$<$	[青蛙] カエルの~
品種	$<$	[金時] サツマイモの~
一品種	$<$	[大納言] アヅキの~
一つ	$<$	[果糖] 果実蜂蜜などに含まれている糖類の~

表 2.2 「生産物」の概念に関する見方の分類の抜粋

	基本的見方
使用に関する見方	使用目的、使用方法、...
製造に関する見方	製造方法、材料、...
構成に関する見方	部分、要素、成分...
性質に関する見方	物理的性質、性能、...
量・単位に関する見方	個数、厚さ、重さ、...
様相に関する見方	色、形、価値、程度、作用相、...
時間に関する見方	存在時間
場所に関する見方	存在場所

第 3 章

WWW からの概念の意味を含む文の抽出

本章では具体的方法について述べていく。WWW からの未知単語の概念の意味を含む文の抽出の流れは図 3.1のようになる。

まず始めに WWW から検索エンジンを使い未知単語を含む文章を抜き出してくる。次に抜きだした文章に対し、機能表現を用いて未知単語の概念の意味を含む文を抽出する。以下の節ではこの一連の流れを具体例を示しながらさらに詳しく説明していく。

3.1 未知単語の決定

未知単語として取り上げたものは、velcro, stapler, paper clip, tablet, セロハンテープ,cellophane tape である。

今回提案した技術では未知単語が複数の意味を持っている時、つまり単語に多義性がある場合にそれを区別する手段を持っていない。したがって機能表現だけではなく多義性の類別ができるような別の方法かあるいは類別できるような機能表現を追加する必要がある。これについては 5 章で述べる。

velcro, stapler, paper clip では、WWW 上での機能表現の存在可能性を調べる。また今回設定した機能表現の分析と社会における単語の認知度の差によって結果がどうなるか調べることも目的としている。tablet は velcro などとの比較研究、さらには多義性のある単語についての抽出の可能性を調べるために未知単語として取り上げた。

セロハンテープ,cellophane tape は日本語による概念抽出が英語の場合とどう異なるのかについて比較、考察を加えるために未知単語として決定した。また

tablet については多義性のある未知単語を実際に抽出した時、どのような結果が得られるのかを調べるために未知単語とした。

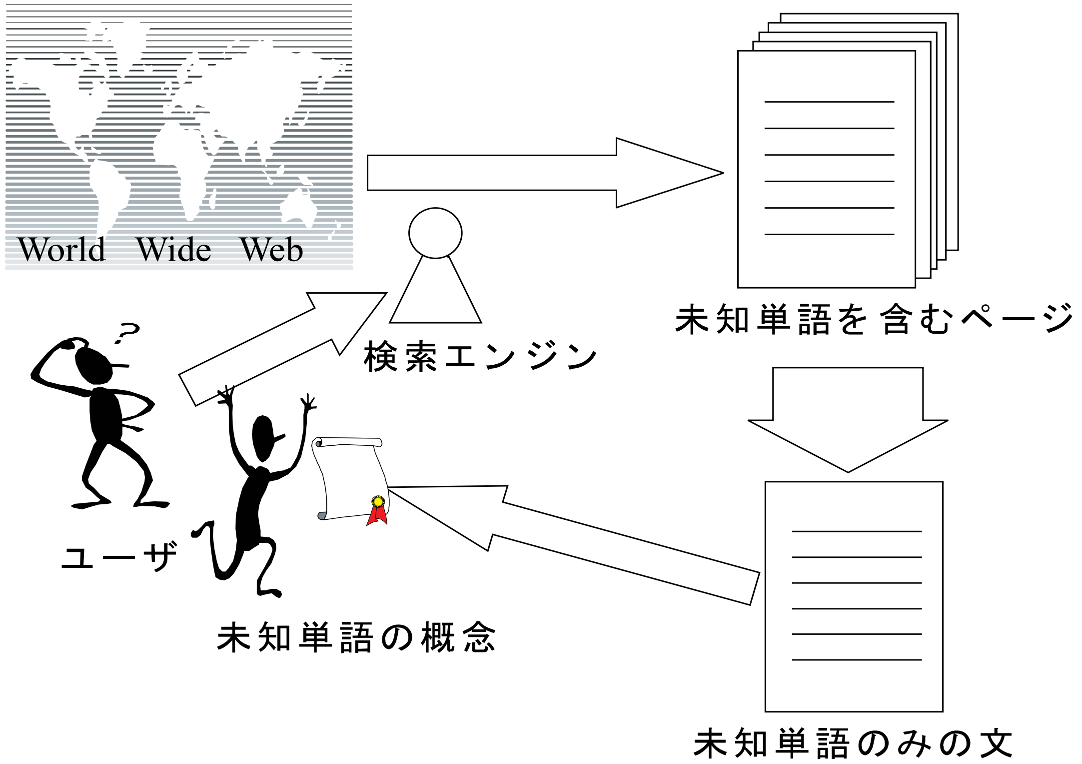

図 3.1 概念抽出の流れ

3.2 WWW からの文章の獲得

WWW からの未知単語を含む文章の獲得は、検索エンジンと httpdown というフリーウェアを利用して行なった。検索エンジンで文章を獲得すると、獲得される文章の質が検索エンジンの能力によって左右される恐れがある。これにより後の概念の抽出の結果に関しても影響を及ぼす可能性があるが、検索エンジンの能力が向上すれば自ずと解消されることであり、本研究で用いた検索エンジンの是非について議論は行なわない。

今回利用した検索エンジンは必ずしも客観的ではないかもしれないが、なるべく多くのデータを確保するため世界最大のデータ数を誇っておりヒット数が多く、さらに語順を保持できるといった理由により goo * を用いることにした。

httpdown は、goo による検索の結果から得られた URL からページを獲得するためのものである。このソフトを使うことにより自動的に大量のページを抜き出すことが可能になる。

さらに獲得したページから未知単語を含む文だけを抽出する。これは UNIX の grep の機能を使って行なった。獲得したページと未知単語を含む文だけを抽出した例を図 3.2と図 3.3に示す。

3.3 未知単語の概念の意味を含む文の抽出

本研究では英語と日本語では抽出方法が異なる。

英語における未知単語の概念の意味を含む文の抽出は、機能表現を利用して行なう。これにより WWW 上の概念の意味を含む文に機能表現が存在するか調べることができる。また今回設定した機能表現を分析し、どのような機能表現が妥当なのかの考察も行なうことができる。未知単語の概念の意味を含む文の具体的な抽出方法は、機能表現が未知単語を含む文に含まれていた場合、その文を抽出し、機能表現の分類にしたがって表示をすることである。図 3.2のページから機能表現を用いて抽出した結果を図 3.4に示す。

日本語については、機能表現による抽出ではなく、人間の手作業による抽出を

*http://www.goo.ne

```
<HTML>
            <HEAD>
            <TITLE>NASA Shuttle Web: Crew Answers Your Questions</TITLE>
            <META NAME="description" CONTENT="Pilot Jim Wetherbee answers a questi
on about Velcro." >
            </HEAD>

            <BODY TEXT="#293D93" BGCOLOR="#FFFFFF" LINK="#ff0000" ALINK="#FFFFFF"
 VLINK="#000088">
            <! - If you edit this map do not forget  menu_sidebar.map for serversi
de->
            <map name="menu_sidebar">

            <!- MAIN BODY OF PAGE ->

           <H1 align="center"><A HREF="index.html">The Crew Answers Internet Quest
ions</A></H1>

David Hopkins of Wardsville, Missouri, asks:<br>
        I saw the shuttle and Mir pass over my position at ~0600 CDT this morning!
 Excellent view.
My question is, I noticed that the Velcro attachment points in the shuttle cabin ar
e both blue and yellow (or orange, it's hard to say).  Any significance to the diff
erent colors?

           <p><A HREF="../wetherbee.html">Jim Wetherbee</A> replies:
           <br><A HREF="answer2.wav">Audio</A>
           <br>
          <table align=right width=170><tr><td align=right><A HREF="/sts-86/images
/esc/Flightday05/lores/S86E5333.jpg"><img src="answer2sm.jpg" width=150 border=0></
A><br clear=all><font size="2" face="arial,helvetica">Jim Wetherbee in Spacehab</fo
nt></td></tr></table>
 <P>
 Velcro, as you know, is a friend of astronauts. It's very easy to lose things up in space.
 We're on a camping trip, but since there's no gravity, things tend to float away, and
 we use Velcro to help keep things in place.
The blue Velcro is permanent Velcro that exists on all of the vehicles in standard locations
and the leave it on the orbiter from flight to flight.
The yellow Velcro is crew option that we can put in locations that make things easier for us, for our
particular mission. Some crews put more  yellow Velcro around the vehicle and multiple crews like
 those locations, and so lately I've noticed that the folks who process the vehicles have been
 leaving some of the yellow Velcro on because it's  useful to multiple crews. Eventually if it
 stays in one place long enough then they change it to blue and it becomes a standard location.
 The other interesting thing to note is that on the structure of the vehicle in the cabin,
 we always put the  pile part of the Velcro, and then all of the objects, the movable objects
 that we move from place to place, like cameras and pens and pencils and pockets, we put
the pile side of the Velcro so that everything is standard.
          <p align=center><b><A HREF="answer3.html">Next Question</A></p>

            <!-- END MAIN BODY OF PAGE -->


             <H5 ALIGN=CENTER><A HREF="http://station.nasa.gov/">International Spa
%ce Station</A><BR><A HREF="/">NASA Shuttle Web</A> | <A HREF="http://www.nasa.gov/"
%>NASA Home</A> | <A HREF="/sts-86/help.html">Help</A></H5>
%
%             <CENTER>
             <ADDRESS> Curator:
             <A HREF="mailto:sarah.enticknap1@jsc.nasa.gov">Sarah Enticknap</A><BR
>
%             Responsible NASA Official: <A HREF="mailto:kelly.o.humphries1@jsc.nas
%a.gov">Kelly Humphries</A><BR>

             Updated:  5 October, 1997
             </ADDRESS>
             </CENTER>
            </td></tr></table>

            </body>
            </html>
```

図 3.2 未知単語 (velcro) を含むページの例

My question is, I noticed that the Velcro attachment points in the shuttle cabin are both blue and yellow (or orange, it's hard to say).
Velcro, as you know, is a friend of astronauts.
We're on a camping trip, but since there's no gravity, things tend to float away, and we use Velcro to help keep things in place.
The blue Velcro is permanent Velcro that exists on all of the vehicles in standard locations and the leave it on the orbiter from flight to flight.
The yellow Velcro is crew option that we can put in locations that make things easier for us, for our particular mission.
Some crews put more yellow Velcro around the vehicle and multiple crews like those locations, and so lately I've noticed that the folks who process the vehicles have been leaving some of the yellow Velcro on because it's useful to multiple crews.
The other interesting thing to note is that on the structure of the vehicle in the cabin, we always put the pile part of the Velcro, and then all of the objects, the movable objects that we move from place to place, like cameras and pens and pencils and pockets, we put the pile side of the Velcro so that everything is standard.

図 3.3 未知単語 (velcro) を含む文の例

We're on a camping trip, but since there's no gravity, things tend to float away, and we use Velcro to help keep things in place.
we use Velcro to help keep things in place.

図 3.4 機能表現による未知単語 (velcro) の概念の意味を含む文の抽出結果の例

行なった。これは日本語と英語で概念の意味をを含む文の違いを調べ、日本語での機能表現の抽出可能性を調べるためである。

3.3.1 英語文からの抽出

英語文に対しては機能表現を使って抽出を行なう。これらの機能表現以外にも機能表現として列挙できるものもあるかも知れない。どれが機能表現として妥当であるのかという議論については、WWW から多くの文章を抜きだし、その中から概念獲得に有用な機能表現を抽出していく作業が欠かせない。しかしそのような作業を行なうには膨大なデータを処理する必要がある。

本研究では機能表現は一般にものを説明する時に使われる単語で共通して用いられるものを設定した。したがって今回の機能表現だけでは充分とはいえないかも知れない。しかしそれらを用いた結果と充分な機能表現で行なった結果を比べても異なるのはその中の機能表現が概念抽出を行なうのに妥当な語であるかというところのみであり、WWW 上に機能表現が存在するかという本研究の目的に

表 3.1 機能表現一覧

機能表現	分類	機能表現	分類
use ○○○ to	使用	cm	量・単位
design	使用	inch	量・単位
○○○ for	使用	gram	量・単位
include	構成	product	製造
contain	構成	construct	製造
consist of	構成	nylon	製造
with ○○○	構成	neoprene	製造
require	性質	cent	様相
allow	性質	shape	様相
make	性質	$	様相
provide	性質		

○○○...未知単語

影響は与えない。

表 3.1に今回設定した機能表現の一覧を示す。これらの機能表現が未知単語を含む文に存在していればその文を概念の意味を含む文として抽出する。表に示されている分類は、表 2.2の「生産物」の概念に関する見方の分類にしたがって独自に分類した。この分類に従うことが妥当であるかについては、議論のあるところである。彼らは「生産物」に対する見方を示しただけであり、概念を知るために有効であるとは限らない。よって分類方法については実験の結果を踏まえた上で後の考察の章で改めて触れる。

ここで機能表現について分類別に説明する。

- 使用に関する機能表現...未知単語の使用方法がわかるような文を抽出できる語を設定した。例えば「 use 未知単語 to ~ 」といった表現が現れれば、その文は未知単語の使用に関して記述されていることがわかる。

- 製造に関する機能表現...未知単語がどういう材料からなるのかを表している文を抜き出せる語とした。今回は nylon と neoprene だけであるが、このような素材に関する語を多く登録することにより、より多くの抽出が可能になる。

- 構成に関する機能表現...未知単語がどこで使われているかに着目をした。つまり未知単語がどの部分で使われているかを表すような語を機能表現とした。

- 量・単位に関する機能表現...ものの大きさや重さを表す語を機能表現とした。cm, gram, inch といった語である。

- 様相に関する機能表現...金額と形を表すものを設定した。

- 性質に関する機能表現...未知単語が備えている機能を示す語を機能表現とした。

3.3.2 日本語文からの抽出

本研究では日本語文からの抽出は手作業で行なう。今回の実験では英語を前提としており、日本語は英語との比較のために実験を行なう。したがって日本語の抽出は人間が単語を理解するのに有用であると思った文を抜きだす。また、抽出した文を用いて日本語での機能表現が存在するか考察を行なう。

第 4 章

実験結果と評価

この章では 3 章で述べた方法によって得られた結果を示す。また 3 章の手順による抽出の結果だけでなく、未知単語を含む文の中の概念を知るのに必要だと思われるすべての文も合わせて示す。これは機能表現による抽出がどれだけ有効なのかを検証するために必要だからである。この抜き出しの作業は人間が手作業で行なった。なおこれらの抽出された文は全て付録に付してある。付録 A は velcro、付録 B は stapler、付録 C は paper clip、付録 D は日本語と英語のセロハンテープ、付録 E は tablet の抽出結果を示している。

また機能表現の抽出結果を評価するために適合率を計算する。適合率とは、抽出した文のうちどれだけ必要な文を抜き出すことができたかを表す数字である。式で表すと、

$$\text{適合率} = \frac{\text{抽出した文の中で必要な文の数}}{\text{抽出した文の数}}$$

となる。この適合率は、機能表現別と機能表現全体と 2 種類表す。これにより、機能表現がどれだけ抽出できたかを示すことができる。さらに今回の機能表現でどれだけ手作業で抽出した文を再現できたかを示す再現率も示す。再現率は、

$$\text{再現率} = \frac{\text{実際に抽出された文の数}}{\text{抽出すべき文の数}}$$

で表される。

以下の節では実験データを示す。velcro 、stapler 、paper clip では、始めに各単語における獲得ページ数、単語を含む文数などを示す。次に未知単語を含む文から機能表現による抽出結果と手作業による抽出結果から、うまく抽出できた例とうまくいかなかった例を示す。

表 4.1 goo の検索結果のスコアの要素

ドキュメントにおける語の頻度	当該ドキュメントのなかに検索語が多く含まれていれば、それだけ高いスコアがつく。
タイトル内の検索語	検索する語をタイトル内に持つページは、検索語を本文にだけしか含まないドキュメントに比べてかなり高いスコアがつけられる。
キーワード内の検索語	「キーワード」META tag のなかに検索する語を含むページは、テキスト内よりも高いランクが与えられるが、タイトル内の語よりも高いスコアにはならない。
ドキュメントの長さ	短いドキュメントのなかに検索語が頻繁に登場する場合、そのページは、長いドキュメントのなかに検索語が登場する場合に比べて高いスコアがつけられる。

cellophane tape では、日本語と英語による抽出結果を示し、日本語と英語各々の文の特徴について触れる。

最後に tablet で多義性のある単語を機能表現で抽出を行ない、単語が複数の意味を持つ場合の抽出結果を示す。

4.1 velcro,stapler,paper clip の概念の意味を含む文の抽出

4.1.1 実験データ

抽出データを表 4.2に示す。ここで goo の検索結果のスコアとは、goo の検索結果に対する信頼度を表現したものである。捜している単語が当該のページに含まれている蓋然性が高ければ、それだけスコアも高いものになっている。各ページのスコアの要素を表 4.1に示す。

velcro は今回実験を行なった単語の中で最もデータが多かった。ページ数では他の単語とそれほど変わらないが、goo のスコアが高く、velcro を含む文が多かった。このことから goo の検索結果のスコアが高ければ、単語の情報量も多いこと

表 4.2 抽出データ

	velcro	stapler	paper clip
ページ数	1200	1100	1100
goo の検索結果のスコア	29755	8382	10618
未知単語を含む文	3048	1223	966
概念の意味を含む文	37	11	12

が伺える。

またページ数の割に velcro を含む文が多いということは、ページ内で velcro が話題の中心となっていることが推測され、概念の意味を含む文が多く存在するものと思われる。

stapler から獲得されたページ数は 1100 と多かったが、velcro に比べ検索結果のスコアと stapler を含む文は大幅に減少している。このことから stapler は velcro ほど話題の中心になっていないことが推測される。

paper clip は stapler と比べると paper clip を含む文数では下回ったが、 goo の検索結果のスコアは上回っている。これらから概念の意味を含む文を stapler より多く抽出できることが期待される。

また図 4.1に抽出した文から筆者が理解した velcro の概念について示す。これから具体的に velcro の形は見えてこないが、どのようなものかおおよその見当がつく。よって抽出結果の表示は単に抽出した文を表示するだけでなく、何らかの方法でより理解ができる表示を行なう必要があるが、それについては第 5 章で述べる。

4.1.2 実験結果

ここでは機能表現による抽出の結果を示す。表 4.3、4.4は機能表現による抽出の適合率と再現率である。これらから単語別で適合率と再現率の違いがあまりない。このことから WWW 上に存在する未知単語を含む文のうち概念の意味を含む文の割合は単語間で大きく異ならないことがわかる。

表 4.5、4.6、4.7は機能表現による抽出の結果得られた文の適合率である。これ

- ものを留めるために使われる
- 簡単に脱着することができる
- 洋服やポケット、バッグ、テニスシューズといった場所で用いられる
- 2cm ~ 5cm 位の大きさ
- 価格は 10.95 ドル

図 4.1 抽出から得られた velcro の概念

らを見てみると機能表現別にうまくいったものといかなかったものに分けられる。使用に関する分類の機能表現は高い適合率であるが、構成や性質に関する分類の機能表現は低い適合率である。ここでなぜ機能表現ごとに適合率が異なるのかを分析する必要がある。そのために分類別に抽出がうまくいった例とうまくいかなかった例を図 4.2と図 4.3に挙げ、その特徴を以下に述べる。

表 4.3 未知単語の機能表現全部による抽出の適合率

	velcro	stapler	paper clip
機能表現による抽出数	75	18	23
概念の意味を含む文の数	29	6	8
適合率 (%)	33.9	33.3	34.8

表 4.4 未知単語の機能表現全部による抽出の再現率

	velcro	stapler	paper clip
手作業による概念の意味を含む文の数	37	11	12
機能表現で抽出できた数	25	5	6
再現率 (%)	67.6	45.5	50

表 4.5 velcro の機能表現別の適合率

機能表現	use	provide	for	with	inch	cm
抽出数	5	1	10	40	2	4
概念の意味を含む文の数	5	1	4	11	0	3
適合率 (%)	100	100	40	27.5	0	75

機能表現	allow	make	design	nylon	neoprene	include
抽出数	1	2	1	5	1	3
概念の意味を含む文の数	1	1	1	1	1	0
適合率 (%)	100	50	100	20	100	0

表 4.6 stapler の機能表現別の適合率

機能表現	use	provide	for	with	make	include	require
抽出数	2	3	3	4	3	1	2
概念の意味を含む文の数	2	1	2	0	0	0	1
適合率%	100	33.3	66.7	0	0	0	50

表 4.7 paper clip の機能表現別の適合率

機能表現	use	inch	require	for	with	contain	gram	make
抽出数	7	4	1	4	2	1	1	3
概念の意味を含む文の数	6	0	0	0	1	0	1	0
適合率%	99.9	0	0	0	50	0	100	0

- 使用の分類の機能表現
  - “Design an object ~” のように「use 未知単語 to」の表現はすべて使用に関してであった。一方、うまくいかなっかた例は “~” のように使用目的の「~のために」ではなく人のためにの「~ために」という意味で使われている文が存在したためにうまくいかなかった。
- 製造の分類の機能表現
  - “Goal Net: ~” のように直接機能表現が未知単語に掛かっている文は概念の意味を含む文として捉えられるが、“A 6cm by 15 cm piece ~” のように直接機能表現が未知単語に掛かっていない文は概念の意味を含む文として捉えられなかった。
- 構成の分類の機能表現
  - うまくいった例、うまくいかなかった例共に「with 未知単語」である。うまくいった例では「~を持っている」という本来意図した意味で使われている文であるが、うまくいかなかった例は「~を使って」というように機能表現が別の意味で使われている文が抽出されている。
- 性質の分類の機能表現
  - 共に provide という機能表現であるが、うまくいった例では未知単語が主語になっているがうまくいかなかった例では未知単語が主語になっていない。この違いによりうまくいく場合とうまくいかない場合に分けられている。
- 量・単位の分類の機能表現
  - これは製造の分類と同じでうまくいったのは直接機能表現が未知単語に掛かっている文で、うまくいかなかったのは機能表現が直接未知単語に掛かっていない文であった。

- 様相の分類の機能表現
  - うまくいった例のみ抽出できた。しかしこれも製造の分類などと同じく、直接機能表現が未知単語に掛かっていなければ抽出はうまくいかないと考えられる。

- 使用の分類の機能表現
  - Design an object that uses Velcro to hold itself together.
- 製造の分類の機能表現
  - Goal Net: At least 20 genuine HT Nylon velcro
- 構成の分類の機能表現
  - Lumbar center cushion with velcro for self-adjustment
- 性質の分類の機能表現
  - Moreover, the stapler provides a removale nonmarring rubber nosepiece
- 量・単位の分類の機能表現
  - One Elastic band 35mm wide 135 cm long with 3 strips of 5cm velcro,
- 様相の分類の機能表現
  - Velcro Gold, chocolate brown, red, blue, green, purple, grey, black white, navy $10.95

図 4.2 うまくいった抽出例

- 使用の分類の機能表現
  - If your hands have taken all they can from pointy ended paper clips, and paper clips that snap together over your fingers, then this is the paper clip for you!
- 製造の分類の機能表現
  - A 6cm by 15 cm piece of 2mm thick leather, with velcro on one side,
- 構成の分類の機能表現
  - Fastened by a zip secured with Velcro.
- 性質の分類の機能表現
  - The materials, plywood, nails, patterns, and stapler are provided by the teacher sponsor and interested students are required to pay enough to cover the cost of these materials.
- 量・単位の分類の機能表現
  - Crumple up a bit of aluminum foil into a neat round ball about a half inch across, and push this onto the part of the paper clip wire protruding from the top of the lid.

図 4.3 うまくいかなかった抽出例

4.2 cellophane tape の概念の意味を含む文の抽出

本節ではセロハンテープという単語を使って、日本語と英語の場合による文章構造の違いを調べる。始めに日本語での抽出の結果の特徴について述べ、その後に英語との比較を行なう。

4.2.1 日本語の実験データ

- 検索結果から獲得したセロハンテープを含むページ数... 255
- セロハンテープを含む文の数... 44
- goo の検索結果のスコア... 255

日本語による検索の結果、抽出された文は検索した中で最少の数だった。英語と比較してもページ数で約 1/4、セロハンテープを含む文にいたっては 1/10 以下である。このような結果になった背景として、次のことがが考えられる。日本語は、世界の中でマイナーな言語である。WWW は世界の人間が集まっているため、公用語としての英語が用いられることが多い。

表 4.8 セロハンテープの抽出の内訳

	使用	量・単位	製造	性質
手作業による抽出数	9	1	1	2

そのような場所では日本語の占める割合は相当少なく、獲得されるデータも大きく差がでてしまう。

さらに検索結果のスコアが 255 と相当低い。これはセロハンテープが話題の中心になっていないし、ページの中でもほとんどでてこないことが伺える。

4.2.2 日本語の実験結果

結果を見てみると、表 4.8のように使用に関する文が全体の 7 割を占めている。残りはセロハンテープの大きさや、セロハンテープの材料に関しての文があるだけであった。この理由は日本語文のデータが少ないためなのか、たまたま使用に関してのみ書いてあったためなのか判断はつかない。これについてはもっとデータを増やす必要がある。

4.2.3 英語の実験データ

- 検索結果から獲得した cellophane tape を含むページ数... 961
- cellophane tape を含む文の数... 466
- goo の検索結果のスコア... 961

英語の実験データは、日本語に比べてデータ量は多い。しかし velcro や stapler などと比べると全てにおいて値が小さくなっている。先の日本語の結果と伴わせて考えると、 cellophane tape という単語は WWW 上ではそれほど多く存在していない可能性がある。もしくは多く存在しているが、検索エンジンの能力の低さ故のことかも知れない。しかし他の単語と比べて cellophane tape だけが著しく低いことから、cellophane tape という単語は WWW 上では多く存在していないと思われる。

よって英語のデータも日本語と同様に概念の意味を含む文はそれほど多くないと考えられる。

4.2.4 実験結果

英語による抽出の結果、日本語に比べて多くの概念の意味を含む文を抽出できた。使用に関する記述は日本語も多かったが、その内容が英語の場合は大きく異なる。図 4.4に抽出例を示すが、日本語の場合、使用に関しては単に「～に貼る」といったものしかなかったが、英語の場合は "So excited, that she covered her hand ～" のようなより使用目的がわかるような記述が見られる。さらに日本語では見られなかった性質に関する記述も見られる。図 4.4の "Many cellophane tapes will take off ～" という記述は cellophane tape の性質を良く表している。また、様相や製造に関しても、""Oh Wow! School Specials Now!" The Los Altos Stationers advertisement ～" のように金額についての文が抽出されたりと日本語の場合と比べて様々な見方の概念の意味を含む文が抽出されている。

何故このような結果になったのか、その理由としては日本語と英語では書かれている文章量が違う、あるいは日本人と他の国の人々では WWW 上で公開する内容が違う等が考えられる。

- So excited, that she covered her hand with cellophane tape to protect it while she washed dishes.
- Many cellophane tapes will take off part of the paper fibers along with
- "Oh Wow! School Specials Now!" The Los Altos Stationers advertisement backed up its headline with such specials as: Scripto ball pens, 7 cents each; cellophane tape, 11 cents; binder paper, 564-count package, $1.16; Flair pens, 37 cents each; all-in-one binder containing note paper, dictionary and theme book index, $1.59.

図 4.4 cellophane tape の抽出例

4.3 tablet の概念の意味を含む文の抽出

本節では多義性のある単語に関して調べる。今回設定した単語は tablet である。tablet には、

1. コンピュータで入力装置として使われるもの
2. 剥ぎ取り式便箋帳
3. 錠剤
4. 鉄道の通票
5. 銘板

- tablet contains 3 mg of melatonin, 10 mg of B-6, in a base of Vitamin A,
- The tablet contains electronics that enable it to detect movement of the cursor or pen and translate the movements into

図 4.5 tablet の抽出例

という意味がある。このように一つの単語で複数の意味が存在する場合、機能表現による抽出の結果は複数の意味を含んだものとなり、本来知りたかった意味で tablet を理解することが困難となる可能性がある。

本節ではこのようなことが実際に起こり得るのか、起こったならばそれはどのような特徴があるのかを述べる。

4.3.1 実験データ

- WWW から獲得した tablet を含むページ数... 約 1100
- tablet を含む文の数... 2674
- goo の検索結果のスコア... 75827

tablet を含む文の数を見ると velcro に次いで高い値を示している。さらに検索結果のスコアが velcro よりも約 3 倍大きい。したがって相当の概念の意味を含む文の抽出が期待できると思われる。

4.3.2 実験結果

抽出結果から概念の意味を含む文のみを見てみると、抽出された文は図 4.5のように 2 つの意味を持っていることがわかる。1 つはコンピュータの入力装置の意味での tablet であり、もう 1 つは錠剤としての tablet である。

抽出された文の内訳を見てみると、文の多くは入力装置の意味での tablet であり、錠剤の意味の tablet は make、 contain といった機能表現の中だけ含まれていた。抽出された文の意味別の割合は、機能表現によって偏っていたり偏って

いなかったりしている。use や inch などでは錠剤の意味が抽出されなかったが、make や contain ではおなじ数だけ抽出されている。これは意味によって一緒に使われる単語が違うためだと思われる。今回の例でいうと、錠剤を表す文では inch や use といった単語は通常一緒に使われないため、これらの機能表現では錠剤の意味の文が抽出されなかったのである。

残りの意味の文は、今回は抽出できなかった。機能表現の数が少なかったため他の意味の文が抽出できなかったのかも知れない。あるいは検索エンジンが他の意味の文を獲得できていなかったのかも知れない。しかしこの実験の主旨は機能表現によって多義性のある語を抽出した場合の特徴を調べることであり、2 つの意味の文が抽出されたことによりその目的は果たされている。

第 5 章

結論・考察

本章では 4 章で得られた結果から導き出される結論、および考察を述べる。本研究により、機能表現を用いた抽出の可能性を調べると共に、WWW 上ではどのような文が多く書かれているのか、またどのような構造で書かれているのかも調べることができた。さらには日本語との比較を行ない、 WWW での日本語の特徴も調べた。

以上のようなことを本章で詳しく述べ、これらをもとにさらに一歩踏み込んだ概念の抽出方法を提案する。

5.1 機能表現による概念の意味を含む文の抽出

今回行なった機能表現による抽出の結果、概念の意味を含む文の抽出に有用な機能表現と有用でないとがはっきりとわかれた。「use 未知単語 to ~ 」と「design 」は全ての未知単語で抽出ができ、その内ほとんどが概念の意味を含む文を抽出している。他方、「product 」と「construct 」は一回も抽出することはなかった。

抽出できた機能表現を見てみると、「 ~に使う 」と「デザインする 」という語である。design は今回偶然このような結果に終わった可能性がある。なぜなら「I designed a pet bottle 」という文が現れる可能性があるからである。 一方「use 未知単語 to ~ 」は使用の機能表現として用いるのは妥当だと思われる。なぜならこの単語は「 ~のために使う 」という使用目的を述べるためにしか用いられないためである。

「make 」や「inch 」などは概念の意味を含む文を抽出できるが、同時に概念を

- velcro
  - Small band width: 1 inch; length: 8 inches; Velcro: 5/8"; Medium band width: 1 inch; Length: 9 inches; Velcro: 3/4 inch; Large band width: 1-1/2 inches; Length: 9 inches; Velcro: 3/4" Made in USA
  - One Elastic band 35mm wide 135 cm long with 3 strips of 5cm velcro,
  - The final part is the ankle holster, a length of velcro 24cm by 5 cm.
  - At the other end loop the elastic is 65cm by 2cm the usual velcro,
- paper clip
  - One manila file folder (11.5 inches by 9 inches) One large (2-inch) paper clip

図 5.1 量・単位についての再抽出

得るには役立たない文まで抽出している。これらの機能表現は条件をきつくすれば精度は上がるものと考える。今回は単に未知単語と機能表現が同じ文中に存在すれば全て抽出した。しかし条件を「length of 未知単語~」という表現が必要とか、「機能表現が直接未知単語にかかっている」ことを条件にしてやれば、抽出される文の数は減るが適合率は上がることが期待される。この条件で量・単位について再抽出を行なうと図 5.1 になる。

また、「with 未知単語」の場合、with が持っている意味が複数あるために意図した通りの文が抽出されなかった。このような多義性のある機能表現は単にパターンマッチを行なって抽出する方法ではうまくいかない。このような多義性のある機能表現では、抽出した文を意味解析する必要がある。

以上のことをまとめると、「use 未知単語 to ~」のように一意に意味が定まる機能表現では概念の意味を含む文を抽出できることがわかった。また一意に意味が定まらない機能表現は抽出が困難であることもわかった。

5.2 日本語コンテンツからの抽出可能性

日本語コンテンツからの抽出を行なう場合、日本語の場合でも機能表現となるものが存在するのか調べる必要がある。それにはまず、日本語文と英語文との違いを見てみる必要がある。

日本語文と英語文でもっとも異なる点は省略された文の存在である。例えば概念を表すために重要な単語が省略されていることに気がつく。例えば次のような文が今回のデータの中で抽出された。

「これの解決は内側へ折り込む長さをけちらないで大胆に張り付けて，ケースの噛み合わせで挟むような気持ちで張り付けるか，セロハンテープなどで補強してください .」

この文からはセロハンテープがどこを補強するのかわからない。どこを補強するのかは、その前の文で書かれていた。またこの文は複数の「張り付ける」という表現が現れていて文の解析が困難である。

このような日本語文の処理は、自然言語の分野の技術を用いて解析する必要がある。

5.3 単語による抽出量の違い

今回の実験では、velcro 、stapler 、paper clip と 3 つの単語を取り上げた。前章でも触れたが、これらの単語の検索結果は各々異なっていた。特に velcro は他の 2 つの単語に比べて多くの文章に含まれており、 goo の検索結果のスコアも高かった。そしてそれはそのまま抽出結果へと反映されている。

単語によって抽出結果の出来が変わってしまう理由は、その単語が社会でどれだけ認識されているかによる。例えば今回の例では velcro の場合のみ概念の意味を含む文が数多く抽出された。このことは、velcro という単語はあまり社会では認識されていないことを表している。なぜなら我々が誰かと話しをしたり、誰かに手紙を書いたりする時、あること (物) について相手が知っていることを知っているならば、我々はそれについて改めて説明することはしない。逆に相手が知らない、あるいは知らないと推測される時、それについて詳しい説明を行なう。

またもう 1 つ考えられる理由は、その単語が話題の中心になりやすいかである。話題の中心になっているということは、その単語についての情報が多く記されることになり、より多くの概念の意味を含む文を抽出できることになる。

```
<B><font color="000080">ONE-WRAP<sup>&reg;</sup> Straps - </font></b><font size="+1
"><a href="ow.html">[Pricing]</a></font> - Are die-cut into re-usable straps made f
rom <font color="red"><B>VELCRO</B></font><tt><sup>&reg;</sup></tt> loop and patent
ed ULTRA-MATE<tt><sup>&reg;</sup></tt> material. Availible in several lengths, widt
hs and 11 different colors for color coding identification
```

図 5.2 構造による抜き出し

5.4 WWW 上の文章の構造

本研究では、国語辞典のように辞書特有の記述方法が取られていない文を対象にしてきた。WWW 上の文は国語辞典の場合のように必ず単文で完結しているとは限らない。したがって国語辞典からの抽出のように一文だけを見ればいいのではなく、その前後の文も見る必要がある。しかし今回のデータを見ると、前後の文脈がなくても概念の意味を含む文が予想以上に抽出された。これは WWW 独自の特徴によるものと思われる。その特徴とは、WWW では文の内容は構造化されていないがその枠組は構造化されていることである。つまり、そのページのタイトルがなんであるのか、そのパラグラフは何について書かれているのかが容易にわかる構造になっている。図 5.2は velcro について書かれている部分である。この図のようにタイトルのような構造化されたものが存在することは、WWW における大きな特徴である。

5.5 概念に関する見方の分類

今回の実験では鶴丸らが提案した分類にしたがって分類を行なった。しかし今回の実験の結果、これらの分類に入らないが、概念の意味を含む文としてみれる表現があった。velcro で分類に入らない概念の意味を含む文を図 5.3に表わす。これらの文は必ずしも未知単語が何であるかについて記されている訳ではない。しかし間接的には未知単語を知ることができ、有用な情報である。したがってこのような情報も載せるような分類を行なうことが必要であると考える。

```
velcro was developed in 1956.
velcro is a registered trademark of Velcro,Inc.
```

図 5.3 分類に入らない概念の意味を含む文

5.6 多義性を持つ単語の抽出

多義性を持つ語の場合、複数の意味が同時に現れて理解を妨げてしまう。ユーザーが全くその単語について未知ならば、意味別に分類してユーザーに提示することが望ましい。

それを実現する方法として、動詞と名詞に着目することが考えられる。何故なら動詞や名詞は特定の名詞と掛かりやすいからである。tablet について見てみると次のような結果が得られる。

- コンピュータの入力装置の tablet ... draw , install , CDROM , DOS 6.2
- 錠剤の tablet ... drink , Vitamin C

これにより分類は可能になるが、このような知識を始めから持っている必要がある。

これに類似した研究として、動詞の語彙的知識獲得における多義性の類別 [9] がある。

5.7 さらに進んだ概念の抽出方法

本節ではいままで本章で述べられたことをもとに、機能表現による抽出だけでなく、その他の方法も取り入れたより精度の高い概念の意味を含む文の抽出方法を提案する。

まず機能表現による抽出であるが、5.1 節でも述べたように一意に意味が定まる機能表現を用いることにより概念の意味を含む文を抽出することが可能である。これによりどのような未知単語がどのようなものか大まかに理解できる。しかし

これだけでは文だけなので、未知単語がどのような形をしているのか具体的に知ることはできない。

そこでもっとも多く文を抽出されたページの URL を提示することを提案する。WWW 上はテキストだけでなく映像も存在するのでユーザに URL を提示することにより、より細かい情報を提供できる。これにより機能表現による抽出を補完することができる。

またその他の方法として、WWW の構造を利用した方法がある。5.4節でも述べたが WWW はページに枠組が存在し、その枠組は概念の抽出に有用であることがわかった。これを使えばより深い理解が可能になる。

さらに多義性の類別は文に現れている動詞や名詞により判断すれば良い。その時、機械を説明する英語 [10] のような類似した動詞をまとめてある本や辞書を利用してやればより精度が上がる。これにより多義性の類別が可能になると思われる。

謝辞

主指導教官である西田豊明教授には、いろいろな面で御指導頂きました。ここに深く感謝致します。

松本裕治教授には、忙しい中、本研究の副指導教官となって頂きました。ここに深く感謝致します。

武田英明助教授には、研究室での活動、学会、研究について御指導頂き、また本研究の副指導教官となって頂きました。ここに深く感謝致します。

上野敦志助手には、論文の紹介や研究について貴重な意見を頂きました。ここに深く感謝致します。

久米出助手には、研究の方向性、研究の進め方など御指導頂きました。ここに深く感謝致します。

研究室の皆様にも大変お世話になりました。

最後に研究室での活動を支えてくれた秘書の小布施文代さんに感謝致します。

参考文献

[1] ISO 704 :Principles and Methods of Terminology,p.16,(1987)

[2] JIS Z 8301 :規格表の様式,p.88,(参考 3:日本工業規格における用語規格のまとめ方)

[3] 梨山 修, 長谷川寿彦 :情報検索システム入門, オーム社,1990

[4] Riichiro Mizoguchi :Task ontology and its use in a task analysis interview system,In *Proceedings of JKAW '92* ,1992

[5] 宇津呂 武仁, 松本裕治, 長尾眞 :日英訳文間の素性構造照合による統語的曖昧性の解消, 情報処理学会論文誌, 第 33 巻, 第 12 号,pp.1555-1564,December 1992

[6] 鶴丸 弘昭, 城戸武雄, 肥後真一, 吉田将 :国語辞典に基づくシソーラスと単語の意味分類に関する一考察, 信学技報 (NLC95-44),Vol95,No.321,pp31-38,1995

[7] 鶴丸 弘昭, 城戸武雄, 日高達, 吉田将 :語義文の機能表現について, 信学技報 (NLC95-65),Vol95,No.429,pp47-54,1995

[8] 鶴丸 弘昭 :語義文からの階層関係の自動抽出, 大規模知識ベースに関する調査研究ーオントロジー工学に関する調査研究ー報告書,p101-125,1997

[9] 内元清貴, 宇津呂武仁, 長尾眞 :動詞の語彙的知識獲得における類義語の用例を用いた多義性の類別, 情報処理学会研究報告,94-NL-101, pp.105-112, May 1994.

[10] 野澤 義延 :機械を説明する英語, 工業調査会,1986.

付録

A. velcro の抽出結果

手作業による抽出

- 使用に関する記述
  - Design an object that uses Velcro to hold itself together.
  - Loop Fastening Systems (velcro)
  - The screen is held in place by Velcro.
  - We use Velcro to help keep things in place.
  - Installs with velcro for fast removal.
  - even necessary as these colorful panels are held in position with Velcro.
  - Velcro and restraining strap designed to secure a notebook computer or other electronic devices.
  - It is Velcro fastened for easy adjustment.
  - The pad is attached to the plastic extension support with Velcro for easy removal.
  - The top pockets are velcro closed for quick access with ane hand.
  - stick on any surface with velcro for easy battery replacement or time adjustment.
  - Lumbar center cushion with velcro for self-adjustment.
- 様相に関する記述
  - It is Velcro fastened. PRICE: $44.95
  - Velcro Gold, chocolate brown, red, blue, green, purple, grey, black white, navy $10.95

- 製造に関する記述
  - High quality velcro material laminated to 1/8 neoprene provides maximum range of adjustment and consistent compression.
  - Custum Goal Net: At least 20 genuine HT Nylon volcro.
- 性質に関する記述
  - The velcro and flap system allows the bag to be ready for quick use.
  - snap or Velcro together with the same ease as the gowns and tops.
  - Quick and easy access to pager is provided by velcro.
  - with velcro so one size fits all men and women.
  - Velcro makes dismantling and reassembling very easy.
  - Attaches easily with Velcro and is comfortable.
  - Heavy-duty metal hardware attachments, velcro
- 量・単位に関する記述
  - A 6cm by 15 cm piece of 2mm thick leather, with velcro on one side.
  - The final part is the ankle holster, a length of velcro 24cm by 5 cm.
  - At the other and loop the elastic is 65cm by 2cm the usual velcro.
  - a hook part of velcro at tips of strap and 2 inch strips of velcro loop on.
- 構成に関する記述
  - Adjustable elastic strap with velcro.
  - Key and Coin Bag with Zipper or Velcro.
  - inner pocket with velcro.

- Velcro (but much better) is attached to the rim of the helmet and the Dule-Lock.
- tennis shoes came out with velcro rather than laces, which worked so much better.
- resistant and attaches to hammock with velcro.
- use clothes with snaps, buttons, or velcro instead of long, loose draw-strings.
- Velcro sealed secure front pockets.
- inner pocket with velcro.

• 場所に関する記述

- has two small electrodes which attach with velcro.

機能表現による抽出

• use

- Design an object that uses Velcro to hold itself together.
- We're on a camping trip, but since there's no gravity, things tend to float away, and we use Velcro to help keep things in place.
- Remember back in the mid-eighties velcro was used to hold everything together, sneakers, jackets, those nylon parachute pants (e.g. Michael Jackson's black & red 'Thriller' outfit).
- I have had a question about using Velcro to line display cases.
- Using Velcro on the flaps to hold it closed and be able to carry it by the

• provide

- Quick and easy access to pager is provided by velcro.

- include
  - In addition, we've included velcro so you can attach the battery pack to your garment of choice for a completely discreet experience!
  - G-String and velcro included!
  - and accessories including velcro.

- velcro for
  - with velcro for easy, compact storage.
  - Use velcro for 100% assurance of no tipovers
  - Extended body with Velcro for quick pull through draw.
  - Installs with velcro for fast removal.
  - with Velcro for easy removal for cleaning.
  - Lumbar center cushion with velcro for self-adjustment
  - Stick on any surface with velcro for easy battery replacement or time adjustment.
  - Auto Charger, and velcro for mounting PDA inside case.
  - velcro, large size clothing, casual clothing for all seasons, maternity wear.
  - Convenient Velcro for easy on/off!

- with velcro
  - The teachers prepare most of the materials with velcro so that the walls and boards are what they call "active walls" or activity oriented.
  - Fastened by a zip secured with Velcro.

- A sheath with velcro on the back.
- elastic with velcro to keep the knife in place.
- The broad, flat shape of the belt clip makes it perfect for attaching to your car dash-board with Velcro or other fasteners.
- problems with the velcro, and especially if you've found a time-tested
- Koala comes with velour leaves and branch which fasten with Velcro to hang from puppeteer's neck.
- Our Wrist Rest Pads have a non-skid bottom and are available with or without a plastic extension support for tight spaces, which can be firmly attached to the desktop with Velcro.
- The wrist and ankle restraints are made of very soft imitation fur and are secured with velcro.
- You owe it to yourself to try this and stop struggling with Velcro and other methods.
- Extended body with Velcro for quick pull through draw.
- fastened to uprights with velcro for ease of set up and
- Eventually, tennis shoes came out with Velcro rather than laces, which worked so much better.
- I was wearing a lightweight outdoor shirt with Velcro
- Take the other length of collar and super-glue it to the other section with the velcro so they are arranged in a T-fashion with the section minus the velcro hanging down. 8
- The VP3 comes with Velcro
- Installs with velcro for fast removal.
- Revolutionary patent-designed partitions connect in seconds with Velcro.

- even necessary as these colorful panels are held in position with Velcro
- Lumbar center cushion with velcro for self-adjustment
- The shoulder guard prevents hair rubs from winter blankets, attaches with velcro, made of strech lycra so not to impede movement in any way.
- resistant and attaches to hammock with velcro.
- with Velcro so one size fits all men and women.
- Stick on any surface with velcro for easy battery replacement or time adjustment.
- Fingertip Timer with Velcro
- Also supplied with Velcro
- Use clothes with snaps, buttons, or velcro instead of long, loose drawstrings.
- receiver loop or on top of a portable and attach the Ferrite Stick with velcro
- Constructed of soft leather with a Velcro
- secure with velcro and other retainers where applicable.
- strong washable plastic with velcro
- pocket with velcro
- has two small electrodes which attach with velcro
- can be had with velcro or snap closure.
- metal hardware, roomy interior pockets with Velcro & zippered
- All openings are fastened with velcro and studs (2" velcro on mast to bow opening, 1" velcro shroud openings)
- A unique "Sandwich-bag roll-foll" secured with velcro, crimps the top down tight, and effectively seals all sizes of pagers in a dry environment.

  - 1 inner pocket with Velcro
  - Attaches easily with Velcro and is comfortable.
  - Elastic Race Wrap with Velcro

- inch
  - a hook part of velcro at tips of strap and 2 inch strips of velcro loop on
  - Small band width: 1 inch; length: 8 inches; Velcro: 5/8"; Medium band width: 1 inch; Length: 9 inches; Velcro: 3/4 inch; Large band width: 1-1/2 inches; Length: 9 inches; Velcro: 3/4" Made in USA

- cm
  - One Elastic band 35mm wide 135 cm long with 3 strips of 5cm velcro,
  - A 6cm by 15 cm piece of 2mm thick leather, with velcro on one side,
  - The final part is the ankle holster, a length of velcro 24cm by 5 cm.
  - At the other end loop the elastic is 65cm by 2cm the usual velcro,

- allow
  - The velcro and flap system allowes the bag to be ready for quick use.

- make
  - Velcro makes dismantling and reassembling very easy.
  - VELCRO Making Puppet Theatre Magic Book and Puppet Set

- design
  - Velcro and restraining strap designed to secure a notebook computer or other electronic devices.

- nylon
  - Velcro fasteners with flexible nylon fittings;
  - Watch band material Nylon and Velcro
  - Goal Net: At least 20 genuine HT Nylon velcro
  - with Velcro and they are attached to webbed nylon strips
  - This 4 piece set is made from nylon with velcro closures.
- neoprene
  - High quality velcro material laminated to 1/8" neoprene provides maximum range of adjustment and consistent compression.

B. stapler の抽出結果

手作業による抽出

- 使用に関する記述
  - Staplers are used for quick fastening jobs.
  - stapler (at home) for thick stack of papers.
  - business presentations, education reports, technological documents and requires only a stapler (concealed) to assemble this unique binder.
  - STAPLERS, PAPER FASTENING; STAPLES AND RELATED ITEMS.
  - The stapler is used to fasten magazine pages together.
- 様相に関する記述
  - Staplers enjoy very profitable careers, you know.
- 製造に関する記述

- Rubber is used on the bottom of the stapler for grip and guard against damage to materials on which it is.

- 性質に関する記述
  - A stapler 5/8 to 1 1/2 and a brad nailer 3/4.
  - the stapler provides a removable nonmarring rubber nosepiece.
  - Most normal staplers are too small to use easily on large booklets, or even if they are big enough, it is difficult to line up the staples properly with the spine of the booklet.

- 場所に関する記述
  - A stapler is available at the stationery table located near.

機能表現による抽出

- use
  - I then (foolishly) decided to use a stapler to do the piercing.
  - The stapler is used to fasten magazine pages together.

- provide
  - There is a stapler provided in the
  - Moreover, the stapler provides a removale nonmarring rubber nosepiece
  - The materials, plywood, nails, patterns, and stapler are provided by the teacher sponsor and interested students are required to pay enough to cover the cost of these materials.

- stapler for

- Rubber is used on the bottom of the stapler for grip and guard against damage to materials on which it is
- Combination paperweight/stapler for Picard's desk
- stapler for convenience stapling •Optional red, blue and

- with stapler
  - menuheadMan who hit woman with ax accused of hitting another with stapler/menuhead
  - Man who hit woman with ax accused of hitting another with stapler
  - unit, 10-bin sorter, 10-bin sorter with stapler, color LCD display, film
  - N32 with stapler and collater

- make
  - The materials used are basic and makes the stapler an inexpensive product.
  - staplers make the coolest noise when you use them.
  - casing of the stapler, because it is easily molded and creates different designs and shapes, making the stapler appealing to the

- include
  - a complete line of pneumatic tools including nailers and staplers; as well

- require
  - This product is ideal for business presentations, education reports, technological documents and requires only a stapler (concealed) to assemble this unique binder.
  - quite a lot of bug fixes... stapler requires to have

C. paper clip の抽出結果

手作業による抽出

- 使用に関する記述
  - A staple or strong paper is sufficient binding.
  - I used a paper clip to secure it to the swing tag.
  - I usually fold the paper over the edge and anchor it with a paper clip.
  - glue place on the cap and paper clip will help tip the position.
  - paper clips of the Gem and Ideal types used to secure paper.
  - Put resume and cover letter in a 9 x 12 envelope and paper clip them together.
  - Do not use ball point pen or paper clip on photos.
  - Use a paper clip to attach any loose barcodes, giftplates, etc.
  - Fasten one end with a paper clip or staple, then unfold the fan.
  - use th epaper clip to attach the cardboard to your kid clothes.
- 性質に関する記述
  - Have each child use a different -style magnet to pick up paper clips.
- 量・単位に関する記述
  - size paper clip has a mass of 0.48 grams.

機能表現による抽出

- use

- That's why the engineers at B & K Paper Clips worked overtime to come up with a paper clip that could be used, like any other, to keep paper together, but that also had the amazing quality of being able to be clipped to your shirt pocket without losing any of its paper-holding-together qualities.
- In this activity, you will use paper clips to make models of different types of
- Reach in and tighten the nuts by hand (with the lockwasher side facing the guitar top) and then use a paper clip, safety pin or small drill bit to tighten the pins and line the holes up parelell to the strings.
- To make a pin, glue the pasta onto a piece of cardboard, then tape a paper clip onto the back–and use the paper clip to attach the cardboard to your kid clothes.
- Use a paper clip to hang the swivel hook from the
- Do not staple or bind the paper–simply use a paper clip to secure the pages.
- Use a paper clip, again, to push down the button inside the little disk ejecting hole, and while holding it down, use a pair of pliers to gently pull the disk loose.

• inch

- One manila file folder (11.5 inches by 9 inches) One large (2-inch) paper clip
- Straighten out the paper clip and cut about 2 inches of the straightened steel
- Crumple up a bit of aluminum foil into a neat round ball about a half inch across, and push this onto the part of the paper clip wire protruding from the top of the lid.

- inches of string and tie a paper clip onto the other end.

- require
  - The SOPS soda machine requires a paper clip, small bits of paper, and an occasional jump start using a knife to run properly.
- paper clip for
  - If your hands have taken all they can from pointy ended paper clips, and paper clips that snap together over your fingers, then this is the paper clip for you!
  - I need a paper clip for the scavenger hunt mom
  - Use thumbtacks or colored map tacks to add eyes and a nose and a paper clip for a wiry tail.
  - Two paper clips for each child
- with paper clip
  - Paper Clip Holder(filled with paper clips)
  - Fare held in place with paper clips.
- contain
  - For example, the first egg holder might contain a paper clip, the second a toy car, the third a grape and so on.
- gram
  - size paper clip has a mass of 0.48 grams.
- make

– These two combinations, make this paper clip the perfect gift for your soar hands which seriously need a vacation from the horrors of everyday paper-handling.

– Stension of water and make a paper clip, that would normally sink, float on the

– Jill said that she could make a paper clip move across her desk without touching the paper clip or the desk.

D. セロハンテープの抽出結果

D.1 日本語文の抽出結果

抽出結果

- 使用に関する記述

  – モーターの軸にきつくセロハンテープを巻いて太くして、自転車用の虫ゴムでつなぎ、さらに虫ゴムをセロハンテープで巻いて固定する。

  – 段ボール箱の内側にアルミホイルを巻き付け、セロテープで固定する。

  – セロハンテープ等でしっかりと貼付してください。

  – ソメイヨシノの花をていねいに分解し、各部分をセロハンテープを使って台紙にはりつけ、各部分の個数を数え、下に記入する。

  – それに、糊代作りセロハンテープを糊代につけてやる。

  – ガラス板をずらし，外径が 8mm のガラス管が入るだけの隙間をあけてセロハンテープで固定した。

  – 接着はセロハンテープで直接紙枠に貼る。

  – ふ入りの葉をろ紙にはさみ、セロハンテープでとめる。

  – かみコップのおしりをくっつけて，セロハンテープでつなぐ。

- 量・単位に関する記述
  - セロハンテープ：幅 18mm 巻芯径 76mm（大巻）
- 製造に関する記述
  - 共重合の有用性を理解するための身近な製品の製造を含む高分子化学実験教材–アクリル酸ブチルの共重合体を用いたセロハンテープの製造
- 性質に関する記述
  - これの解決は内側へ折り込む長さをけちらないで大胆に張り付けて，ケースの噛み合わせで挟むような気持ちで張り付けるか，セロハンテープなどで補強してください．
  - セロハンテープで合わせ目をとめる場合、折り線とテープの線を平行にそろえた方がきれいに仕上がる。

D.2 英語文の抽出結果

抽出結果

- 使用に関する記述
  - So excited, that she covered her hand with cellophane tape to protect it while she washed dishes.
  - But during renovations at some of the district's older schools, workers cut computer trunk lines and, in one instance, used cellophane tape to patch them, district officials said.
  - Don't use cellophane tape to seal frozen foods.
  - Cut the cording (use cellophane tape to keep the cord from unraveling) and cover the wire.
  - Please do not use scotch or cellophane tape to put up signs and posters.

- assembled and held in place with cellophane tape.
- The cellophane tape securing them to the wall had grown old and sticky, and in frustration Boaz leaned his copy of the I Ching against them. He turned off the lamp and leaned back to see if they could still be read.
- But during renovations at some of the district's older schools, workers cut computer trunk lines and, in one instance, used cellophane tape to patch them, district officials said.
- For this procedure, you will need a small piece of masking or cellophane tape to mark the new EEPROM device as blank.
- Put a piece of masking tape or cellophane tape over drywall plaster before nailing to avoid plaster from chipping.
- they had attached it to the paper with cellophane tape.
- The cellophane tape used to piece together an early 20th-century map is now dark and stiff.

• 様相に関する記述

- "Oh Wow! School Specials Now!" The Los Altos Stationers advertisement backed up its headline with such specials as: Scripto ball pens, 7 cents each; cellophane tape, 11 cents; binder paper, 564-count package, $1.16; Flair pens, 37 cents each; all-in-one binder containing note paper, dictionary and theme book index, $1.59.

• 製造に関する記述

- LePage's manufactured a cellophane tape made from wood pulp and adhesive material called "Biodegradable Transparent Tape."

• 性質に関する記述

- I would suggest placing the phone list and any special coupons in a hermetically sealed (cellophane tape is certainly the poor man's option) waterproof envelope.
- Many cellophane tapes will take off part of the paper fibers along with
- Cellophane tape sticks to almost anything and hence is also called "adhesive tape."
- The occasional tape break was easily fixed with some cellophane tape and a pencil to repunch the holes, but mangled tape (if you sat or stepped on a roll) was ugly.
- Do not use aerosol spray glues, cellophane tape, or rubber cement, as they will deterioate rapidly.
- A brass plate is affixed with cellophane tape to the front of the base, waiting for a name to be carved into it.

• 量・単位に関する記述

- Now, take two pieces of cellophane tape each about 4 cm long, and
- Heavy-duty gummed tape or cellophane tape at least 1 1/2 to 2 inches wide

• 場所に関する記述

- Cellophane tape, Post-it notes, overhead transparency film, or anything that sticks to the Graffiti area will work as well.

E. tablet の抽出結果

機能表現による抽出

• use

- Pressure Pen: Assign - Use tablet standard interface should be deselected.
- I want to use a tablet, but HOW can I do that with out a CRASH ?
- for those who use a tablet for drawing
- On the other hand, to use the tablet with all applications, you must install a "mouse" driver that cannot provide access to precise absolute positioning, pressure input, or other advanced features.

- include
  - Tablet/Slip case include

- contain
  - tablet contains 3 mg of melatonin, 10 mg of B-6, in a base of Vitamin A,
  - The tablet contains electronics that enable it to detect movement of the cursor or pen and translate the movements into

- with tablet
  - You can contact Hitachi, and if you send them the original CDROM in its case, they will send out an older version of Dabbler that will work with the tablet.
  - have been a problem with the tablet itself and I sent it back to WACOM and they replaced it with a new one.
  - Has anyone else noticed this to occur with tablets?
  - of reading the documentation that comes with the tablet, I couldn't figure

  - A 105-MB PCMCIA hard drive, DOS 6.2, and Windows for Pen Computing are included with the tablet.
  - Pressure Tool with tablet
  - Only the two players with the tablets share the allied victory
  - The requirement your computer needs to work with the tablet successfully are
  - You will be amazed what you can do with a tablet that you can't

- inch
  - portion of the tablet's 18x24-inch active area as your workspace and
  - to obscure some part of the tablet's whopping 24x30-inch surface with

- gram
  - GDU (Gelatin Dissolving Units) per tablet from 1200 GDU per gram

- allow
  - of com ports or IRQ's would allow the tablet to function at all with this

- make
  - A unique compound of special Imedeen marine extracts, Vitamin C and Zinc are what lets this little tablet make such a big difference.
  - and make the tablets to fall down like cascade.
  - Plus 70 years of experience, with all types of material has enabled us to fine tune our plant, which now enables us to make tablets where others have failed.

– We can't promise we can make your tablets, we can't promise we can make that delivery tomorrow but we can promise that we will keep our word.