

## クロスリンガル情報検索結果の閲覧支援のための 主要キーワード対訳表示の効果

鈴木 雅実 井ノ上 直己 橋本 和夫

KDD 研究所

〒 356-8502 上福岡市大原 2-1-15

E-mail: {msuzuki, inoue, kh}@lab.kdd.co.jp

あらまし 言語間にまたがるクロスリンガル情報検索において、検索結果一覧の段階でコンテンツの取捨選択を容易にするための閲覧支援情報として、各文書中の主要キーワードを利用者の言語に翻訳して提供する枠組を考案した。この目的での閲覧支援の効果は、提供する対訳情報の質に依存すると考えられる。そこで、英語コンテンツの検索結果から文書の選択を支援する提供情報として、(1) 英語の主要キーワードリスト (出現頻度順)、(2) 対訳辞書の第1訳語を用いて英日変換した日本語キーワードリスト、(3) 対訳辞書の訳語候補同士の共起頻度分布を用いて英日変換した日本語キーワードリストを比較評価する実験を行なった。この結果、上記(1),(2),(3)の順に、選択閲覧支援効果が上昇する傾向が見られた。

キーワード 多言語情報検索, クロスリンガル, 閲覧支援, 翻訳, 訳語選択

## Effect on Displaying Translated Major Keywords of Contents as Browsing Support in Cross-Language Information Retrieval

Masami SUZUKI Naomi INOUE Kazuo HASHIMOTO

KDD Research and Development Laboratories

(2-1-15, Ohara, Kamifukuoka-shi, 356-8502 JAPAN)

E-mail: {msuzuki, inoue, kh}@lab.kdd.co.jp

Abstract In a prototype of cross-language information retrieval system, we proposed a framework of providing translated major keywords of contents in the user's language as browsing support information, which can help filtering documents in a hit lists. The effect of providing such browsing support may depend the quality of translation. Thus, in this report, we describe the comparison through an experiment, where the following different conditions were examined for giving browsing support in a retrieval list: 1)Frequent (original) English keywords list for each document, 2)Translated Japanese keywords list using a default translation in the lexicon, 3)Translated Japanese keywords list generated in referring a cooccurrence distribution among the translation candidates within Japanese corpus. As a result, an increase of support effect was observed in conditions 1) through 3).

key words cross-language information retrieval, translation, browsing support

## 1 はじめに

筆者らは、これまでの報告等で、言語間にまたがるクロスリンガル情報検索<sup>1</sup>における特有の問題として、検索結果一覧からの文書の取捨選択に関わる閲覧支援情報の重要性を指摘している。また、この点を実践的に検証するため、試作した多言語情報検索サーバ CLINKS の検索閲覧支援機能として、検索結果一覧中の各文書から抽出した主要キーワードのリストを、利用者の言語に翻訳して提示する枠組を提案した [6]。さらに、キーワードのリストをより適切に翻訳するため、翻訳生成側の言語のコーパス内での訳語候補同士の共起分布に基づいて訳語の組み合わせを選択する方法を組み込む等の工夫を行なっている [7] [8]。本稿では、このような閲覧支援情報の効果を測定するために行なった評価実験について述べ、その結果および関連する問題を考察する。

## 2 クロスリンガル情報検索における閲覧支援の問題

### 2.1 閲覧支援の意義

筆者らが試作した多言語情報検索サーバ CLINKS では、利用者の検索要求(複数入力を可とするキーワード)とは異なる言語の検索結果が導かれた際に、まず表示される検索結果一覧において、該当する各文書の主要キーワードのリストが利用者の言語で提供される。ここで、閲覧支援情報の意義について述べる。

通常の単一言語内に閉じた(モノリンガル)情報検索においては、検索結果一覧において文書のタイトル、URL、文書サイズ、更新年月日等の文書属性情報のほかに、コンテンツの一部のテキストが表示されることが多い。これは、文書の選択閲覧、すなわち検索結果を利用した情報のナビゲーションに役立つ支援情報として、検索サーバ側から提供されるものである。WWW のコンテンツを対象とするサーチエンジンの場合では、コンテンツの最初の数行程度のテキスト情報がしばしば用いられるが、文書の特徴を適切に反映したものとなっているかについては疑問が残る。他の支援情報の提示方法としては、コンテンツ中の主要なキーワードや関連性のあ

るキーワードをリストアップする手法が用いられる場合が、一部のサーチエンジンで見られる。

一方、クロスリンガル情報検索においては、検索対象言語に精通した利用者でない限り、コンテンツの一部が原語で表示されたとしても、文書の取捨選択に直接役立つことは相当に難しい(あるいは多くの時間を要する)と考えられる。従って、文書の選択閲覧を支援する情報は利用者の言語に翻訳されて提供されることが望ましい。しかし、現状の機械翻訳技術に基づいてテキストの一部の翻訳情報を生成した場合や、前述した、主要キーワードのリストを翻訳情報として提供する場合(CLINKS)は、その翻訳/訳語の正確性が問題となる。しかしながら、多少の不正確性があっても翻訳情報として提供されることによる利用者へのメリットが生ずるものと考えられる。

### 2.2 クロスリンガル情報検索支援の関連研究

クロスリンガル情報検索に関連する研究開発事例は、最近になって増加してきているが、検索内容に関して何らかの利用者の(検索要求)言語による付加情報を与える閲覧支援に関する事例は比較的少ないのが現状である。その中で幾つかの代表例を記す。

#### • TITAN [5]

検索結果一覧中に英日翻訳された HTML 文書のタイトルが表示される。ただし、HTML 文書のタイトルは文書の内容を代表するには非常に一般的であったり(欠落している場合もある)、便宜的に付与されることも多く、翻訳が成功したとしてもコンテンツの探索に十分有効とは限らないのが問題であろう。なお、TITAN の最新版では、検索結果を文書の類似度計算によってクラスタリングした結果を表示する機能を設けており、検索結果の一覧を支援する一つの方法として注目すべき試みと言える。

#### • QUILT [3]

英語の検索要求に対しスペイン語の検索結果が導かれるが、検索結果一覧中には文書(記事)番号が示されるのみで、そこから選択された文書については、原文書(スペイン語)とその逐語訳文書(英語)が提示される。これに関連して、機械翻訳の枠組が利用されてはいるものの、検

<sup>1</sup>cross-language information retrieval, translingual text retrieval

索結果一覧の段階では文書のタイトルと原語テキストの一部が表示され、利用者の要求に応じて翻訳が起動される検索サービスの代表例として、Altavista が挙げられる。

● (Resnik 97) [9]

この研究事例では、WWW コンテンツのうち1件が数行程度のイエローページ<sup>2</sup>を対象として、多言語 (翻訳) の簡約 (“gisting”) を生成するプロトタイプが紹介されている。その目的は、元のコンテンツに対して、“*decision-making*” を可能とする最小限の情報を提供しようとすることであり、gisting の手法は、それが *decision-making* にどれだけ役立つかによって評価されるべきだとしている。Resnik が主張する評価基準は、次のような点を判断する材料を与えるか否かである。

- そのリンクをたどる価値がある。
- テキストの一定部分を翻訳する価値がある。

文献 [9] では、この考えに従った評価実験が報告されている。原文のテキストに対して、名詞部分の辞書引き結果が示された “gisting” を参照して、被験者が各イエローページ項目を予め指定された6種類のジャンルに分類するというタスクを実行した結果を示し、この種の情報でも利用者の判断を支援する上で十分役立つと結論づけている。

これらの参考事例に示すように、検索結果の一覧中からの所望のコンテンツの探索に関する限り、現在の閲覧支援方式はまだ十分であるとは言えない。Resnik [9] の問題意識には筆者らと共通するものがあるが、そこで報告された事例は、限定された領域の極めて短いテキスト・コンテンツを対象としていることにより、通常の辞書引き程度の情報 (可能性のある訳語を列記したもの) でも、有用な場合があることを示すに留まっており、クロスリンガル情報検索における、種々の支援情報の提供方法の有効性の比較については未検証であると言える。より効果的な検索支援システムを設計する上での指針を得るためには、様々な閲覧支援情報の提供方法を比較検討する必要がある。ここで、現在試みられている方法について、提供情報の形式/内容の組み合わせによる分類を行ない、それぞれの利害得失を記したものを表1に掲げる。

この表1に示すように、支援情報の提供方法については様々な形態が考えられるが、情報抽出上の問題 (要約/キーワード抽出等) と翻訳生成上の問題が絡み合うため、支援効果における優劣を直接比較評価することは難しい。そこで、次章で述べる実験報告では、まず、筆者らの多言語検索サーバ CLINKS における、出現頻度を反映した主要キーワードのリストを翻訳する枠組の有効性を検証するための評価を行ない、その結果に基づいて、さらに他の手法 (要約テキストの提示等) との比較方法を検討して行くこととした。

表 1 選択閲覧支援のための翻訳情報の提示方法と問題点

提示形式	生成上の問題	支援効果/問題点
部分テキスト (最初の数文または一定文字数)	文の抽出は容易。翻訳ソフトの利用が可能。	翻訳による品質の劣化が問題。また、最初に重要な情報が含まれるとは限らない。
部分テキスト (自動要約)	要約/翻訳ソフトの利用。検索要求に応じて、高速に実行可能であることが必要。	要約/翻訳による品質の劣化が問題。前項も含め、クロスリンガル情報検索支援に使われた例はないと思われる。
主要キーワードのリスト (頻度を反映)	原語でのリストの生成は比較的容易。翻訳生成方法が問題。	翻訳品質の差が影響すると思われる。(CLINKS の例に相当)
主要キーワードのリスト (出現順)	生成が比較的容易。	同上。テキスト最初部分の逐語訳に近い。(Resnik の例にほぼ該当する。)
文書のタイトルの翻訳	翻訳ソフトの利用	タイトル情報の欠落/不適切性、翻訳の質の問題。

<sup>2</sup>一種のリンク見出しのようなもので、企業名とその所在地、サービス内容を簡潔に記した内容

従って、本稿の範囲では、表1のうち、コンテンツ内の出現頻度に基づく主要キーワードのリストの提示による選択閲覧支援効果について、キーワードの表示内容(生成方法)の違いが実際にどの程度影響を及ぼすかを測定するために行なった評価実験を中心に報告する。

### 3 閲覧支援機能の評価実験

ここで述べる評価実験は、クロスリンガル情報検索の利用者が、母国語と異なる言語の文書検索結果が一覧表示される段階で、種々の形式で与えられる各文書の主要キーワード・リストを参照して、探索課題に該当する文書をどの程度正しく選択できるかを比較しようとするものである。比較条件は、キーワードの表示言語(英語/日本語)、翻訳(日本語)キーワードの場合の生成方法(訳語品質)の違い、キーワード表示個数、被験者の対象言語能力等である。以下、実験方法とその結果について順に記す。

#### 3.1 実験の被験者

日本人の大学生を主体とする19~32歳の男女64名の協力を得た。

#### 3.2 実験対象と探索課題

評価実験用に用意したテキスト・コーパスは、中国の経済分野の新聞記事(原文は中国語)で、これを英語および日本語に翻訳したものである<sup>3</sup>。記事数は224件で、これを用いて8組の探索課題と、それに対する仮想的な検索結果を作成した。各探索課題は、同コーパスの日本語版の記事タイトルから選んだ。すなわち、被験者に対して、「中国の小売業の現状」のような探索テーマを与えると同時に、該当記事を含む10件の記事検索結果を用意した。各探索課題に対する10件の検索結果集合については、実際にキーワード検索を実行するのではなく、224件の記事データに対し、コンテンツ中のキーワード分布の類似した記事をクラスタリングする処理を行ない、類似度がある程度近い同一クラスタに分類された記事10件づつを用いた[1][2]。探索課題のテーマ一覧は、表2に示す通りである。

表2 探索課題のテーマ一覧

1.	中国の小売商業の現状
2.	広東省は住宅改革の新構想を出す
3.	インフラ建設における海外企業の協力期待
4.	対外経済貿易の新構想 - 戦略的産業の技術輸入が優遇
5.	アジアは鋼材消費が増加する傾向
6.	昨年中国の食糧輸入が大幅に降下
7.	我が国のセメント規格改正
8.	農村のバイク市場潜在力は大きい

#### 3.3 実験手順

##### 1. 被験者へのガイダンス

被験者は、最初に実験に関する説明を記した文書を読んだ後、英語能力に関する筆記テストを受けた(時間制限10分、100点満点)。続いて、モニタースクリーン上での操作説明を聞いた後、課題の実行に取りかかった。

##### 2. 課題の実行

被験者は8個の探索課題を順次実行する。各課題を選択すると、ブラウザの画面上部にその探索テーマが表示され、下方には10件の検索結果を含む一覧が表示される(スクロール無し)。個々の文書の検索結果として、文書番号(直接のコンテンツ情報は持たない)とともに示される、キーワードリストを参考情報として、被験者は、該当テーマに近いと思われる文書の番号を、指定された入力欄に上位3位まで記入する。確認ボタンをクリックした時点で、次の課題に移行する。同様に、8個すべての課題を実行して終了する。

#### 3.4 主要キーワードの提示条件

被験者が探索課題における該当文書の選択のために参照する、各文書の主要キーワードの提示条件は次の通りである。

##### 1. 表示内容

- 条件A: 原語(英語) キーワード
  - 条件B: 翻訳(日本語) キーワード
- 英語キーワード群に対して、各々の英日対訳辞書中の最初の訳語を当てたもの

<sup>3</sup>英語版の1記事当たりの語数は300~500語程度である。

- 条件 C：翻訳（日本語）キーワード

英語キーワード群に対して、各々の英日対訳辞書中の訳語候補の中から、日本語コーパス内の共起分布を参照して、尤度の高い組み合わせを選択したもの<sup>4</sup> [8]

## 2. 表示個数

- 3個
- 6個
- 9個
- 12個

なお、各被験者に対する実際の提示方法は、上記表示内容／個数と探索課題の組み合わせにより、8種類のパターンを用意した。付録1に被験者に提示される探索課題の実行画面の例を示す<sup>5</sup>。

## 3.5 実験結果

以下の記述において、「正解率」とは、各探索課題において被験者が選択した文書番号（最大3個まで）が、実際の探索テーマが示す記事と一致した度合を指すものとする。

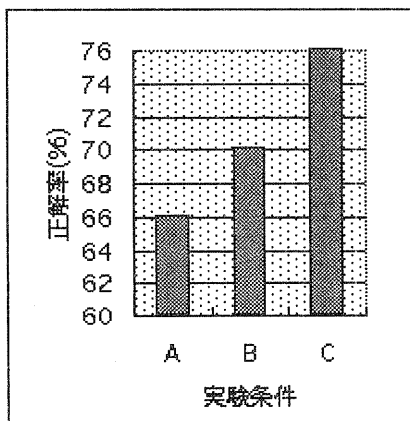


図1 キーワード表示条件と正解率

<sup>4</sup>ここで用いた日本語コーパスは、インターネット上から収集したWWWコンテンツで、種々のニュース情報、技術解説、アジア圏の文化紹介等を主体とした、約7,000ページ相当のHTMLである。

<sup>5</sup>通常の検索結果表示では、文書タイトル（この場合記事の表題）が示されるのが普通であるが、本実験では主要キーワード自体の支援効果を測定するため、付録1のような提示方法を採用した。

### 3.5.1 主要キーワードの表示内容と正解率

キーワードの表示内容の提示条件 A, B, C の各場合に、被験者全体の正解率として得られた値を下図の図1に示す。

また、各課題毎に正解文書の選択のし易さは異なることが予想されるが、次の表3に、正解率の数値に従って課題全体を2分した場合の条件別の正解率を示す。さらに、被験者が各探索課題への回答に費やした平均の時間について、条件 A, B, C を比較した結果を表4に示す。

表3 探索課題と正解率

課題番号	正解率			
	全体	条件 A	条件 B	条件 C
3	50 %	48 %	53 %	62 %
1	53 %			
6	56 %			
2	64 %			
5	81 %	85 %	85 %	89 %
7	81 %			
4	87 %			
8	89 %			

表4 探索課題の回答所要時間

	条件 A	条件 B	条件 C
所要時間	145 sec.	104 sec.	105 sec.

### 実験結果による観察事項

図1の全体の平均を示す正解率の比較により、条件 A, B, C の順で、被験者が該当文書を正しく選択した割合が増加したことが分かる。また、探索課題を低正解率群と高正解率群に分けて比較すると、低正解率群においてこの傾向が顕著であり、主要キーワードの対訳情報（およびその質的向上）の効果が鮮明に現れているものと考えられる。さらに、1探索課題における回答までの平均所要時間は、キーワードが英語で表示される場合と、日本語で表示される場合とでは前者が後者の約1.4倍程度の長さとなっており、大きな差が見られる。

### 3.5.2 主要キーワードの表示個数と正解率

表5に、主要キーワードの表示個数を変化させた場合の各正解率を掲げる。これを見る限り、キーワード表示個数と正解率の間にはほとんど相関が無いことが分かる。

表5 キーワード表示個数と正解率

	3個	6個	9個	12個
低正解率群	58%	38%	63%	60%
高正解率群	94%	85%	85%	83%

### 3.6 被験者の個人属性と正解率との関係

被験者に対して、実験の前に英語能力を判定するための筆記試験<sup>6</sup>を実施した。また、実験後に被験者のアンケートを取り、インターネット／情報検索等への親近性を聴取した。これらの個人属性と正解率の関係を、参考のため表6、表7に示す。

表6 英語筆記テスト得点と正解率

英語得点分類	低(24名)	中(30名)	高(10名)
正解率	65%	73%	69%

(注) 英語テスト得点による分類基準は次の通り。

低: 49点以下, 中: 50~69点, 高: 70点以上。

表7 インターネット等への親近性と正解率

属性分類	1(17名)	2(17名)	3(15名)	4(15名)
正解率	68%	71%	67%	75%

(注) 属性グループの分類は次の通り。

1. パソコン, インターネットとは無縁。
2. 大学等でインターネット経験が少しある。
3. インターネット, パソコンに触る機会が多い。
4. インターネット, 情報処理に相当習熟。

表6および表7の集計結果は、今回の実験で使用した英語能力測定テスト, あるいはインターネット等への親近性と, 正解率との間で相関があると裏付けるような示唆を与えるものではない。

### 3.7 実験結果の総合的な分析

今回の評価実験の目的は、クロスリンガル検索結果の一覧表示の段階で、コンテンツの主要キーワードのリストが表示された場合、その表示内容の言語(コンテンツの原語あるいは利用者の言語)の違いや翻訳の質、キーワードの表示個数等が、探索テーマに適合する文書を推定する際にどのように異なる支援効果をもたらすかを検証しようとするものである。3.5および3.6に示した実験結果から、この実験の範囲内で次のような現象が確認された。

<sup>6</sup>英語能力に関してある一面を測定するもので、大規模な能力試験とは異なる。

## 1. 主要キーワードの表示内容／個数による支援効果の相違

被験者にとっての外国語である英語で表示された場合に比して、被験者の母国語である日本語で表示された場合の方が、検索結果一覧の段階で目的に合致するコンテンツを選択することが容易であり、判断に要した時間も短縮された。また、同様に日本語で表示された場合でも、英語のキーワードを基に辞書引きを行ない最初の訳語を用いる方法と、翻訳生成側の日本語コーパス内の訳語候補同士の共起分布を考慮して訳語を生成する方法とでは、後者の方が、コンテンツの選択をより容易にした。さらに、以上のような支援効果の程度は、コンテンツ選択の難度が高い場合により顕著であった。一方、キーワードの表示個数と支援効果との間には、明確な相関関係は見られなかった。

## 2. 被験者の属性と実験結果との関係

今回の実験の参考情報として被験者から得た、(一面的な)英語能力, およびインターネット等の経験度の相違が、実験結果に影響を与える可能性については、確定的な判断はできないが、ほぼ無視してもよいと考えられる。

## 4 今後の検討課題

本稿の評価実験の対象は、内容の分野および形式としては均一のコンテンツを対象としたものであったが、WWWの多種多様なコンテンツを対象とする場合は、ある程度の量のテキスト(たとえば、通常のブラウザ設定でスクロール無しで視認できる範囲で、一まとまりの整合性を持ったドキュメント)を対象とした、支援方法を確立する必要がある。この条件では、2章の表1で挙げた要約手法や、今回の評価実験対象とした、主要なキーワードの対訳表示等の試行が、さらに精度を向上させることにより、利用可能な技術となることが期待される。特に、筆者らの方法では、適切な訳語候補を含む対訳辞書の整備と、検索対象(たとえば英語コンテンツ)と分野的に近い、翻訳生成言語(日本語)の十分なテキストコーパスの確保が前提条件となる。

また、対象とするコンテンツの分野、テキストのスタイル／量、利用者の言語能力等の様々な要因が介在する点に注意を払う必要がある。支援効果へ

の影響度が低い要素を予備実験の段階で確認した上で、共通の基盤で閲覧支援方法間の比較が行なえるような実験計画が求められる。この場合、比較尺度は、本報告で用いたような、ある課題に対する正解率や所要時間等の複数の視点を設定するほか、同時に表示可能な情報の量等を考慮した比較方法も検討に値する。たとえば、(翻訳)要約テキストを表示する方法では、コンテンツ1件当たりのテキスト量が増加するほど読解に時間を要すると思われるが、取捨選択のための判断情報量が単調に増加するかどうかは疑問である。今回の実験では、キーワード表示個数とコンテンツ選択の正解率の相関は不明であったが、テキスト表示の場合の様相も、実験的に確認する必要がある。

さらに、クロスリンガル情報における閲覧支援の問題を考える上では、コンテンツ作成者の協力も今後の情報共有の多言語化・多角化を進展させる上で、欠かせない要素となってくることが予想される。すなわち、発信するコンテンツが他の言語の読者の目に触れることを意識すれば、発信者側からの協力手段として、現状の枠組内で用意されているHTMLの機能を用いて、他言語によるアクセスを支援するための工夫が可能であり、現時点の検索/翻訳等の自然言語処理関連技術の応用範囲が広がることになる。たとえば、コンテンツで使用される言語/コードセットの記述すら無いHTML文書も多いが、このようなレベルから、コンテンツ作成者側が陽に指定するキーワードやサマリー等を、タグを用いて記述しておくようなレベルまで、コンテンツに関する知識(メタ情報)が埋め込まれていれば、検索サーバが提供する閲覧支援情報を生成する上で助けとなる。この点については、機械翻訳やWWWの国際標準化等の研究開発者の立場からの提言もなされている[4]。

## 5 おわりに

クロスリンガル情報検索における検索結果の選択閲覧支援に関する問題は、本稿で指摘した通り、まだ十分な検討が行なわれていない分野で、今後さらに試行を必要とするものである。多種多様なコンテンツの増大により、その中から所望の情報を発見することがますます困難になって行くことが予想される現在、情報源と利用者を仲介する役割として検索サーバに求められる機能は、コンテンツの輪郭を

簡潔に示す、一種の情報「編集」であると考えられる。この観点から、要約手法等も含む、コンテンツの重要部分の抽出と、翻訳技術の連携を、言語をまたがる情報検索をより現実的なものとするための重要課題としたい。


## 謝辞

本研究の遂行に際してご指導、ご助言を頂いたKDD研究所の村谷所長・村上前所長・鈴木副所長に感謝いたします。また、有益なコメントを寄せられた関連研究グループの諸氏に深謝いたします。さらに、システム試作の過程および評価実験の実施に際してご協力頂いた社内外の方々に、この場で厚くお礼申し上げます。


## 参考文献

- [1] 青木 圭子, 松本 一則, 橋本 和夫, “類似ドキュメントの発見手法の検討”, 情報処理学会第54回全国大会, 3-39, 1997.
- [2] 青木 圭子, 松本 一則, 橋本 和夫, “大量文書向けのクラスタリング手法の評価”, 情報処理学会第56回全国大会, 1998.
- [3] M. W. Davis and W. C. Ogden: “Implementing Cross-Language Text Retrieval Systems for Large-scale Text Collections and the World Wide Web”, *AAAI Spring Symposium Cross-Language Text and Speech Retrieval Electronic Working Notes*, 1997.
- [4] NetWorld Interop 98: “Conference on Multilingualis and the Internet”, presentation material, 1998.
- [5] 菊井 玄一郎, 他: “インターネット情報ナビゲーションにおける多言語機能”, 自然言語処理の応用に関するシンポジウム, 情報処理学会, pp.97-106, 1995.
- [6] M. Suzuki and K. Hashimoto: “Enhancing Source Text for WWW Distribution”, *Proc. of Workshop on Information Retrieval with Oriental Languages (IROL'96)*, pp.51-56, Taejon, 1996.
- [7] 鈴木 雅実, 井ノ上 直己, 橋本 和夫: “多言語情報検索における利用者支援について - 主要キーワードの対訳付与に関する検討 -”, 情報処理学会自然言語処理研究会, NL122-11, 1997.
- [8] 鈴木 雅実, 井ノ上 直己, 橋本 和夫: “クロスリンガル情報検索における閲覧支援機能について”, 1998.
- [9] P. Resnik: “Evaluating Multilingual Gisting of Web Pages”, *AAAI Spring Symposium Cross-Language Text and Speech Retrieval Electronic Working Notes*, 1997.

(この例の場合は翻訳されたキーワードを6個表示)

 検索結果一覧 (HK006)

検索項目:[我が国のセメント規格改正]

第一選択  第二選択  第三選択   (入力したら次へ)

- (1) 0693 /CLINKS/WWW/english/news/newsDB/1997/05/23/news03.txt  
(中国語, 食品および薬品管理, 草本論, 製品, 米国, パーセント)
- (2) 0401 /CLINKS/WWW/english/news/newsDB/1997/04/07/news01.txt  
(セメント, 体積, 意見, パーセント, 建設, 輸送)
- (3) 0311 /CLINKS/WWW/english/news/newsDB/1997/03/24/news01.txt  
(セメント, 市場, 年間, 住居, 中国, 国内)
- (4) 0722 /CLINKS/WWW/english/news/newsDB/1997/05/28/news02.txt  
(セメント, 市場, 年間, パーセント, 正方形, 元)
- (5) 0711 /CLINKS/WWW/english/news/newsDB/1997/05/27/news01.txt  
(資料, 年間, 料金, 請求, 耕すこと, 農業機械)
- (6) 0643 /CLINKS/WWW/english/news/newsDB/1997/05/16/news03.txt  
(モトローラ, 上海, 冒険, 共同, 機械, 会社)
- (7) 0393 /CLINKS/WWW/english/news/newsDB/1997/04/04/news03.txt  
(ビーバー, ヒマシ油, 世界, 中国, 開発, その日)
- (8) 0723 /CLINKS/WWW/english/news/newsDB/1997/05/28/news03.txt  
(セメント, 植物, 料金, トン, 体積, 年間)
- (9) 0083 /CLINKS/WWW/english/news/newsDB/1997/02/18/news03.txt  
(織物, 綿, 中国, 生産, 連邦化, 証明書)
- (10) 0682 /CLINKS/WWW/english/news/newsDB/1997/05/22/news02.txt  
(セメント, 国際規格, 基準, 国民, 産業, 国)