Ad hoc IR task

Group's ID:

BKYTR

List of Run ID(s):

1. BKJJBIDS
2. BKJJBIFU
3. BKJJDCFU

* NTCIR-1 = NACSIS Test Collection 1

1 Indexing

1.1 Indexing

1) Index units for Japanese text (uni-gram, bi-gram, other n-gram, word, phrase, other):

BKJJBIDS: bi-gram (description only run)
BKJJBIFU: bi-gram
BKJJDCFU: word, phrase

2) Index units for English text (n-gram, word, phrase, other):

word

3) Index units for English terms within sentence in Japanese!Jn-gram, word, phrase, other):

word

4) Method(s) used in indexing (lexicon, morphological analysis, other):

BKJJDCFU: Longest-match against a lexicon was used to segmented Japanese text into words.

5) Method(s) used in selection of index terms (e.g.: stop word, type of characters, part of speech, etc.):

All hiragana characters were discarded in indexing.

6) Standardizing terms (characters)?: NO

7) Stemming Algorithm?: NO

8) Term Weighting?: YES

9) Phrase identification?: NO
- Kind of phrase:
- Method used (statistical, syntactic, other):

10) Syntactic Parsing?: NO

11) Thesaurus and/or lexical resources?: YES

The dictionary in Chasen and a small Japanese/English dictionary

12) Word sense disambiguation?: NO

13) Spelling checking (including manual checking)?: NO

14) Correcting them?: NO

15) Proper noun identification?: NO

16) Method(s) used in tokenizing?:

BKJJBIDS, BKJJBIFU: An adjacent pair of kanji or katakana is treated as one token.
BKJJDCFU: The longest-match against a dictionary is used to segment Japanese text into tokens.

17) Use "YOMI" of Japanese text?: NO

18) (if 17 is Yes) method(s) used to generate "YOMI":

19) Other method(s) used in indexing (Please specify):

1.2 Index data structures built from NTCIR-1

1) Kind of index structures
- Inverted index: YES
- Clusters:
- Signature files:
- Pat-tree:
- Knowledge bases:
- Other(s) (Please specify):

2) Summary of index
- Run ID: BKJJBIDS
- Total storage [in MB]: 250
- Total time to build [in hours]: 1/2 hours
- Automatic process? (If not, number of manual hours [in hours]): YES
- Use of positional information (off-set)?: NO

- Run ID: BKJJBIFU
- Total storage [in MB]: 250
- Total time to build [in hours]: 1/2 hours
- Automatic process? (If not, number of manual hours [in hours]): YES
- Use of positional information (off-set)?: NO

- Run ID: BKJJDCFU
- Total storage [in MB]: 160
- Total time to build [in hours]: 1/3 hours
- Automatic process? (If not, number of manual hours [in hours]): YES
- Use of positional information (off-set)?: NO

1.3 Data built from sources other than NTCIR-1

1) Internally-built auxiliary files
- Domain:
- Type of file (thesaurus, knowledge base, lexicon, etc.):

Lexicon

- Type of representation: Flat text file
- Total computer time to build [in hours]: 1
- Total manual time to build [in hours]: 0
- Total manual time to modify NTCIR-1 (if already built) [in hours]: 0
- Use of manual labor?: NO

2) Externally-built auxiliary files (including commodities):

Part of our dictionary was derived from the dictionary in Chasen and Edict (a small Japanese/English dictionary).

2 Query construction

2.1 Automatically constructed queries

1) Average computer time to build query [in CPU seconds]: Approx. 0.5

2) Method(s) used in building queries
   - Tokenizing (uni-gram, bi-gram, n-gram, word, phrase, other):

     BKJJBIDS, BKJJBIFU: bi-gram
     BKJJDCFU: word

   - Phrase extraction from topics?: NO
   - Syntactic parsing?: NO
   - Word sense disambiguation?: NO
   - Proper noun identification?: NO
   - Automatic expansion of queries?: NO
     * Preciously-constructed tools such as thesaurus?:
     * Automatic relevance feedback?:
       + Local context analysis:
       + Other(s) (Please specify):
     * Other(s) (Please specify):
   - Automatic addition of Boolean/procimity operators?: NO
   - Other(s) (Please specify):

2.2 Manually constructed queries: NONE

1) Average time to construct query [in minutes]:

2) Person constructing queries
   - Domain expert:
   - Computer system expert:
   - Other (Please specify):

3) Tools used in building query
   - Word frequency list?:
   - Knowledge base?:
   - Other lexical tools (e.g. thesaurus, lexicon)? (Please specify):

4) Method used in query construction
   - Term weighting?:
   - Boolean operators (AND, OR, NOT)?:
   - Proximity operators?:
   - Addition of terms not included in topics?:
     * Source of additional terms:
   - Other (Please specify):

2.3 Interactive queries: NONE

1) Initial query constructed automatically or manually:

2) Person doing interaction
   - Domain expert:
   - Computer system expert:
   - Other (Please specify):

3) Average time to complete an interaction [in minutes]):

4) What determines the end of an interaction:

5) Method(s) used in interaction:
   - Automatic term reweighting from relevant documents (relevance feedback)?:
   - Query expansion from relevant documents (relevance feedback)?:
     * All terms in relevant documents added:
     * Only top X terms added (specify X):
     * User selected terms added:
   - Query modification in manual?:

3 Searching

3.1 Search times

1) Run ID: BKJJBIDS
2) Computer time to search [average per query, in CPU seconds]: 1.58302

1) Run ID: BKJJBIFU

2) Computer time to search [average per query, in CPU seconds]: 3.02057

1) Run ID: BKJJDCFU

2) Computer time to search [average per query, in CPU seconds]: 2.16566

3.2 Searching methods

1) Vector space model?:

2) Probabilistic model?: YES

3) Other (Please specify):

3.3 Factors in ranking

1) TF (Term Frequency)?: YES

2) IDF (Inverse document frequency)?: YES

3) Other term weights? (Please specify):

4) Semantic closeness?: NO

5) Positional information in the document?: NO

6) Syntactic clues? NO

7) Proximity of terms?: NO

8) Document length?: YES

9) Other (Please specify):

Query length, collection length.

3.4 Machine information

1) Machine type for the experiment: PC (Intel Pentium II)

2) Was the machine dedicated or shared?: shared

3) Amount of hard disk storage [in MB]: 9000

4) Amount of RAM [in MB]: 256

5) Clock rate of CPU [in MHz]: 450

3.5 Others

1) Brief description of features of your system not answered by above questions:

2) Others (Please specify):

3)Your group has:
- Japanese native speaker(s): one
- Member(s) who can understand Japanese language: YES
- No member who can understand Japanese language: NO

随時検索タスク（Ad hoc IR task）

チーム略称：CRL

実行ID（複数ある場合はすべて）：CRL3 CRL4 CRL5 CRL6 CRL7 CRL8 CRL9 CRL10 CRL12 CRL13 CRL14

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：なし

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：なし

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：なし

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：なし

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：なし

(6)語彙（文字）の正規化を行なったか？：いない

(7)ステミングアルゴリズムを用いたか？：使った

(8)語の重みづけを用いたか？：用いた

(9)フレーズ単位で索引づけをしたか？：用いた
・フレーズの種類は？：すべて
・フレーズの見つけ方は？（統計的、構文的、その他）:その他

(10)構文解析は行なったか？：いない

(11)シソーラスや用語集などを用いたか？：用いた

(12)語義の曖昧性解消は行なったか？：していない

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：していない

(14)誤字脱字やスペルの修正は行なったか？：していない

(15)固有名詞を識別したか？：していない

(16)どのような方法で索引単位に分割したか？：その他

(17) 日本語のヨミを用いたか。用いていない

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：なし

(2)索引の概要
・実行ID：
・索引の規模［MB］：
・構築に要した時間［時間］：
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：
・語の出現位置（オフセット）は使用したか？：

全システムとも索引を使用せず

１．３　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：該当なし
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：EDR

２　検索式の作成

２．１　自動的に作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：60

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語とフレーズ
・フレーズの抽出：形態素解析
・構文解析：使用せず
・語義の曖昧性解消：せず
・固有名詞の識別：せず
・検索式の自動拡張：
－シソーラスなど既存のツール：　せず
－自動レレバンスフィードバック：せず
＊ローカルコンテクストアナリシス：せず
＊その他（具体的に）：せず
－その他（具体的に）：CRL4 のみはje0のデータから獲得した類義語辞書で自動拡張
・ブール演算子や近接演算子などの自動的付与：
・その他（具体的に）：

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：15

(2)誰が検索式を作成したか？
・分野の専門家：一部の分野のみ
・計算機システムの専門家：せず
・その他（具体的に）：せず

(3)検索式作成に用いたツール
・語の出現頻度リスト：用いた
・知識ベース：用いた
・その他の辞書的ツール（シソーラスや辞書など具体的に）：EDR

(4)検索式作成に用いた方法
・語の重みづけ：用いた
・ブール演算子(AND, OR, NOT)：せず
・近接演算子：せず
・検索課題に含まれていない語の追加：せず
－追加した語の情報源：せず
・その他（具体的に）：せず

２．３　対話的な検索式の作成

対話的な検索せず

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（1課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位X個の検索語を追加（Xはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

3　検索

3．1　検索時間

(1)実行ID：CRL3,4,6,9 は，

(2)検索時間（1検索式に対する平均CPU時間［秒］）：1800

その他は 900

3．2　検索モデル

(1)ベクトル空間型を用いたか？：せず

(2)確率型を用いたか？：?

(3)その他（具体的に）：ロバートソン式を簡略化したもの

3．3　ランクづけの要素

(1)TF（語の出現頻度）：使った

(2)IDF：使った

(3)その他の重みづけ（具体的に）：質問文間のIDF

(4)意味の近さ：使用せず

(5)文書中の位置：使用せず

(6)構文的な手がかり：使用せず

(7)語の近接（距離）：使用せず

(8)文書の長さ：使用せず

(9)その他（具体的に）：せず

3．4　計算機についての情報

(1)実験に使用した計算機：Sun Ultra 10

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量［GB］：60GB

(4)RAMの総容量［MB］：1GB

(5)CPUのクロック数［MHz］：?

3．5　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

もう，ご存知かもしれませんが，
trec のツールでは正解の記事番号が複数あると
それだけ分正解にしてしまうようです．
評価に用いられる際には注意して下さい．

(3)チームの構成員に：
・日本語のnative speakerがいる：いる
・日本語のわかる人がいる：いる
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称：CRL

実行ID（複数ある場合はすべて）：CRL15

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：なし

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：なし

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：なし

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：なし

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：なし

(6)語彙（文字）の正規化を行なったか？：いない

(7)ステミングアルゴリズムを用いたか？：使った

(8)語の重みづけを用いたか？：用いた

(9)フレーズ単位で索引づけをしたか？：用いた
・フレーズの種類は？：すべて
・フレーズの見つけ方は？（統計的、構文的、その他）:その他

(10)構文解析は行なったか？：いない

(11)シソーラスや用語集などを用いたか？：用いた

(12)語義の曖昧性解消は行なったか？：していない

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：していない

(14)誤字脱字やスペルの修正は行なったか？：していない

(15)固有名詞を識別したか？：していない

(16)どのような方法で索引単位に分割したか？：その他

(17) 日本語のヨミを用いたか。用いていない

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：なし

(2)索引の概要
・実行ID：
・索引の規模［MB］：
・構築に要した時間［時間］：
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：
・語の出現位置（オフセット）は使用したか？：

全システムとも索引を使用せず

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：EDR

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：60

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語とフレーズ
・フレーズの抽出：形態素解析
・構文解析：使用せず
・語義の曖昧性解消：せず
・固有名詞の識別：せず
・検索式の自動拡張：
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：CRL4 のみはje0のデータから獲得した類義語辞書で自動拡張
・ブール演算子や近接演算子などの自動的付与：
・その他（具体的に）：

2．2　手動で作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均時間［分］）：15

(2)誰が検索式を作成したか？
・分野の専門家：一部の分野のみ
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：用いた
・知識ベース：用いた
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：用いた
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

2．3　対話的な検索式の作成

対話的な検索せず

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：CRL15

(2)検索時間（１検索式に対する平均CPU時間［秒］）：500

３．２　検索モデル

(1)ベクトル空間型を用いたか？：

(2)確率型を用いたか？：

(3)その他（具体的に）：ロバートソン式を簡略化したもの

３．３　ランクづけの要素

(1)TF（語の出現頻度）：使った

(2)IDF：使った

(3)その他の重みづけ（具体的に）：

(4)意味の近さ：使用せず

(5)文書中の位置：使用せず

(6)構文的な手がかり：使用せず

(7)語の近接（距離）：使用せず

(8)文書の長さ：使用せず

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：Sun Ultra 10

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量［GB］：60GB

(4)RAMの総容量［MB］：1GB

(5)CPUのクロック数［MHz］：?

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：いる
・日本語のわかる人がいる：いる
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称： DOVE

実行ID（複数ある場合はすべて）： DOVE1, DOVE2, DOVE3

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：ＮＡ

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：品詞、字種、ストップワード

(6)語彙（文字）の正規化を行なったか？：２バイト→１バイト変換

(7)ステミングアルゴリズムを用いたか？：用いていない。

(8)語の重みづけを用いたか？：用いた

(9)フレーズ単位で索引づけをしたか？：いえ。
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：いえ。

(11)シソーラスや用語集などを用いたか？：いえ。

(12)語義の曖昧性解消は行なったか？：いえ。

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：いえ。

(14)誤字脱字やスペルの修正は行なったか？：いえ。

(15)固有名詞を識別したか？：いえ。

(16)どのような方法で索引単位に分割したか？：形態素解析プログラムを用いた。

(17) 日本語のヨミを用いたか。：いえ。

(18) ヨミを用いた場合、ヨミはどのように生成したか。：ＮＡ

(19)索引づけに用いたその他の方法（具体的に）：特になし。（形態素解析結果を品詞、字種、ストップワードでフィルタリング）

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引： Yes
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：
・索引の規模 [MB] ：180 MB
・構築に要した時間［時間］：約１日
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：いる。
・語の出現位置（オフセット）は使用したか？：いえ。

１．３　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル：なし
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量 [MB] ：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：なし

２　検索式の作成

２．１　自動的に作成した検索式 DOVE1（DOVE2, DOVE3 も DOVE1 の結果を初期状態としてインタラクションを開始）

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：約 3 ms（形態素解析と品詞等によるフィルタリングのみ）

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語
・フレーズの抽出： No
・構文解析：No
・語義の曖昧性解消：No
・固有名詞の識別： No
・検索式の自動拡張： No
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与： No
・その他（具体的に）： No

２．２　手動で作成した検索式 No

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成　　DOVE2 と DOVE3

(1)最初の検索式の作成は自動的か手動か：自動

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：参加者

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：４～５分

(4)検索を終了した理由は何か？：原則として１フィードバック

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：DOVE2
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：DOVE2 ２００個
－ユーザが選択した語を追加：DOVE3
・手動での検索式の修正を行なったか？：いえ

3 検索

３．１ 検索時間

(1)実行ID：

(2)検索時間（１検索式に対する平均CPU時間［秒］）：１～２秒程度

３．２ 検索モデル

(1)ベクトル空間型を用いたか？：Yes

(2)確率型を用いたか？：

(3)その他（具体的に）：

３．３ ランクづけの要素

(1)TF（語の出現頻度）：Yes

(2)IDF：Yes

(3)その他の重みづけ（具体的に）：

(4)意味の近さ：No

(5)文書中の位置：No

(6)構文的な手がかり：No

(7)語の近接（距離）：No

(8)文書の長さ：Yes

(9)その他（具体的に）：

３．４ 計算機についての情報

(1)実験に使用した計算機：Dec Alpha Server 8200

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量 [GB] ：40GB

(4)RAMの総容量 [MB] ：1000 MB

(5)CPUのクロック数 [MHz] ：300 MHz

３．５ その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：Yes
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称： JALAB

実行ID（複数ある場合はすべて）：JALAB

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：該当データなし

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：該当データなし

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：その他

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：

(6)語彙（文字）の正規化を行なったか？：該当データなし

(7)ステミングアルゴリズムを用いたか？：該当データなし

(8)語の重みづけを用いたか？：該当データなし

(9)フレーズ単位で索引づけをしたか？：該当データなし
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：該当データなし

(11)シソーラスや用語集などを用いたか？：該当データなし

(12)語義の曖昧性解消は行なったか？：該当データなし

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：該当データなし

(14)誤字脱字やスペルの修正は行なったか？：該当データなし

(15)固有名詞を識別したか？：該当データなし

(16)どのような方法で索引単位に分割したか？：

(17) 日本語のヨミを用いたか。：該当データなし

(18) ヨミを用いた場合、ヨミはどのように生成したか。：該当データなし

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：JALAB
・索引の規模［MB］：143
・構築に要した時間［時間］：6
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：自動化
・語の出現位置（オフセット）は使用したか？：未使用

１．３　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：該当データなし
・ファイルの型（シソーラス、知識ベース、辞書など）：該当データなし
・総記憶量［MB］：該当データなし
・表現された概念数：該当データなし
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：該当データなし
・構築に要した手作業の時間［時間］：該当データなし
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：該当データなし
・手作業を行なったか？：該当データなし

(2)外部で構築された補助ファイル（商品含む）：該当データなし

２　検索式の作成

２．１　自動的に作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：該当データなし

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：該当データなし
・フレーズの抽出：該当データなし
・構文解析：該当データなし
・語義の曖昧性解消：該当データなし
・固有名詞の識別：該当データなし
・検索式の自動拡張：該当データなし
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：該当データなし
・その他（具体的に）：該当データなし

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：該当データなし

(2)誰が検索式を作成したか？：該当データなし
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール：該当データなし
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法：該当データなし
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：該当データなし
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：該当データなし

(2)誰が検索を実行したか？：該当データなし
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：該当データなし

(4)検索を終了した理由は何か？：該当データなし

(5)対話で使用される方法：該当データなし
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：JALAB

(2)検索時間（１検索式に対する平均CPU時間［秒］）：21

３．２　検索モデル

(1)ベクトル空間型を用いたか？：○

(2)確率型を用いたか？：

(3)その他（具体的に）：

３．３　ランクづけの要素

(1)TF（語の出現頻度）：

(2)IDF：○

(3)その他の重みづけ（具体的に）：

(4)意味の近さ：

(5)文書中の位置：

(6)構文的な手がかり：

(7)語の近接（距離）：

(8)文書の長さ：

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：IBM PC365(Pentium Pro)

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量 [GB] ：2.1

(4)RAMの総容量 [MB] ：64

(5)CPUのクロック数 [MHz] ：200

３．５　その他

(1)上の質問で回答していないシステムの特色：検索の高速性

(2)その他（具体的に）：該当データなし

(3)チームの構成員に：
・日本語のnative speakerがいる：○
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称： JSCB

実行ID（複数ある場合はすべて）：jscb1,jscb2,jscb3

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語、フレーズ

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：ストップワード、品詞

(6)語彙（文字）の正規化を行なったか？：はい

(7)ステミングアルゴリズムを用いたか？：いいえ

(8)語の重みづけを用いたか？：はい

(9)フレーズ単位で索引づけをしたか？：はい
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）:構文的

(10)構文解析は行なったか？：いいえ

(11)シソーラスや用語集などを用いたか？：いいえ

(12)語義の曖昧性解消は行なったか？：いいえ

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：いいえ

(14)誤字脱字やスペルの修正は行なったか？：いいえ

(15)固有名詞を識別したか？：いいえ

(16)どのような方法で索引単位に分割したか？：形態素解析

(17) 日本語のヨミを用いたか。：いいえ

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置ファイル：はい
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：JSCB1,JSCB2,JSCB3
・索引の規模［MB］：228
・構築に要した時間［時間］：1.7
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：はい
・語の出現位置（オフセット）は使用したか？：

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：

2　検索式の作成

2．1　自動的に作成した検索式:jscb1,jscb2

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語、フレーズ
・フレーズの抽出：はい
・構文解析：いいえ
・語義の曖昧性解消：いいえ
・固有名詞の識別：いいえ
・検索式の自動拡張：はい（JSCB1,JSCB2）
－シソーラスなど既存のツール：いいえ
－自動レレバンスフィードバック：はい
＊ローカルコンテクストアナリシス：いいえ
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：いいえ
・その他（具体的に）：

2．2　手動で作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

2．3　対話的な検索式の作成:jscb3

(1)最初の検索式の作成は自動的か手動か：自動

(2)誰が検索を実行したか？：
・分野の専門家：

・計算機システムの専門家：はい
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：１７分２６秒

(4)検索を終了した理由は何か？：時間制限（２０分）

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：はい
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：はい
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：閾値でカットオフ
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：いいえ

３　検索

３．１　検索時間

(1)実行ID：JSCB1,JSCB2

(2)検索時間（１検索式に対する平均CPU時間［秒］）：10.6（１検索式に対する検索式自動作成も含む実時間[秒]）

３．２　検索モデル

(1)ベクトル空間型を用いたか？：はい

(2)確率型を用いたか？：いいえ

(3)その他（具体的に）：

３．３　ランクづけの要素

(1)TF（語の出現頻度）：はい

(2)IDF：はい

(3)その他の重みづけ（具体的に）：

(4)意味の近さ：

(5)文書中の位置：

(6)構文的な手がかり：

(7)語の近接（距離）：

(8)文書の長さ：

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：DellOptiPlex GX1

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量 [GB] ：9

(4)RAMの総容量 [MB] ：384

(5)CPUのクロック数 [MHz] ：450

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：はい
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称： K3200

実行ID（複数ある場合はすべて）： K32001,K32002

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）： その他のn-gram

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）： n-gram

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）： n-gram

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：その他

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：選択なし(全文字列を索引語に)

(6)語彙（文字）の正規化を行なったか？： バイト文字を対応するバイト文字に変換した。

(7)ステミングアルゴリズムを用いたか？： 用いない。

(8)語の重みづけを用いたか？： 用いない。

(9)フレーズ単位で索引づけをしたか？： しない。
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？： しない。

(11)シソーラスや用語集などを用いたか？： 用いない。

(12)語義の曖昧性解消は行なったか？：行なわない。

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：行なわない。

(14)誤字脱字やスペルの修正は行なったか？：行なわない。

(15)固有名詞を識別したか？：しない。

(16)どのような方法で索引単位に分割したか？： 字種ごとにn-gramに分割

(17) 日本語のヨミを用いたか。： 用いない。

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：B-Tree

(2)索引の概要
・実行ID：K32001
・索引の規模［MB］： -
・構築に要した時間［時間］： -
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：自動化されている
・語の出現位置（オフセット）は使用したか？：使用した

・実行ID：K32002
・索引の規模［MB］： -
・構築に要した時間［時間］： -
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：自動化されている
・語の出現位置（オフセット）は使用したか？：使用した

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル：使用していない。
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：使用していない。

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）： -

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語
・フレーズの抽出： しない。
・構文解析： しない。
・語義の曖昧性解消：しない。
・固有名詞の識別：しない
・検索式の自動拡張：しない。
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：検索単語をOR結合
・その他（具体的に）：

2．2　手動で作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：K32001
(2)検索時間（１検索式に対する平均CPU時間［秒］）：-
(1)実行ID：K32002
(2)検索時間（１検索式に対する平均CPU時間［秒］）：-

３．２　検索モデル

(1)ベクトル空間型を用いたか？：用いた。

(2)確率型を用いたか？：用いない。

(3)その他（具体的に）：

３．３　ランクづけの要素

(1)TF（語の出現頻度）：利用した。

(2)IDF：利用した。

(3)その他の重みづけ（具体的に）：利用しない。

(4)意味の近さ：利用しない。

(5)文書中の位置：利用しない。

(6)構文的な手がかり：利用しない。

(7)語の近接（距離）：利用しない。

(8)文書の長さ：利用した。

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：UNIXマシン

(2)その計算機は専用か共用か：-

(3)ハードディスクの総容量 [GB] ：-

(4)RAMの総容量 [MB] ：-

(5)CPUのクロック数 [MHz] ：-

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：います。
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）
チーム略称： KCTRG
実行ID（複数ある場合はすべて）：KCTRG1: 検索要求を利用
KCTRG2: 検索要求、検索要求説明を利用
※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ
１．１　索引づけに用いた方法
(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語
(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：
(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語
(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析
(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：品詞
(6)語彙（文字）の正規化を行なったか？：NO
(7)ステミングアルゴリズムを用いたか？：NO
(8)語の重みづけを用いたか？：YES
(9)フレーズ単位で索引づけをしたか？：NO
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：
(10)構文解析は行なったか？：NO
(11)シソーラスや用語集などを用いたか？：NO
(12)語義の曖昧性解消は行なったか？：NO
(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：NO
(14)誤字脱字やスペルの修正は行なったか？：NO
(15)固有名詞を識別したか？：NO
(16)どのような方法で索引単位に分割したか？：形態素解析
(17) 日本語のヨミを用いたか。：NO
(18) ヨミを用いた場合、ヨミはどのように生成したか。
(19)索引づけに用いたその他の方法（具体的に）：単語間共起頻度、ＴＦ・ＩＤＦ

１．２　NTCIR-1から構築された索引データの構造
(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：○
・その他（具体的に）：使用していない
(2)索引の概要
・実行ID：
・索引の規模［MB］：
・構築に要した時間［時間］：
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：YES
・語の出現位置（オフセット）は使用したか？：NO

１．３　NTCIR-1以外の情報源から構築されたデータ：未使用
(1)独自に構築した補助ファイル
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？
(2)外部で構築された補助ファイル（商品含む）：

２　検索式の作成
２．１　自動的に作成した検索式
(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：
(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語
・フレーズの抽出：NO
・構文解析：NO
・語義の曖昧性解消：NO
・固有名詞の識別：NO
・検索式の自動拡張：NO
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：NO
・その他（具体的に）：

２．２　手動で作成した検索式：行なっていない
(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：
(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：
(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：
(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成：行なっていない
(1)最初の検索式の作成は自動的か手動か：
(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：
(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：
(4)検索を終了した理由は何か？：
(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索
３．１　検索時間：プロトタイプ版であるので発表できません。
(1)実行ID：
(2)検索時間（１検索式に対する平均CPU時間［秒］）：

３．２　検索モデル
(1)ベクトル空間型を用いたか？：YES
(2)確率型を用いたか？：NO
(3)その他（具体的に）：

３．３　ランクづけの要素
(1)TF（語の出現頻度）：○
(2)IDF：○
(3)その他の重みづけ（具体的に）：単語共起頻度
(4)意味の近さ：×
(5)文書中の位置：×
(6)構文的な手がかり：×
(7)語の近接（距離）：文内一律
(8)文書の長さ：×
(9)その他（具体的に）：

３．４　計算機についての情報
(1)実験に使用した計算機：Sun Ultra 30

(2)その計算機は専用か共用か：共用
(3)ハードディスクの総容量 [GB] ：9
(4)RAMの総容量 [MB] ：256
(5)CPUのクロック数 [MHz] ：296

３．５　その他
(1)上の質問で回答していないシステムの特色：
(2)その他（具体的に）：
(3)チームの構成員に：
・日本語のnative speakerがいる：YES
・日本語のわかる人がいる：YES
・日本語のわかる人はいない：NO

随時検索タスク（Ad hoc IR task）

チーム略称：KLAB

実行ID（複数ある場合はすべて）：KLAB1,KLAB2

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：文字列照合

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：名詞

(6)語彙（文字）の正規化を行なったか？：NO

(7)ステミングアルゴリズムを用いたか？：NO

(8)語の重みづけを用いたか？：NO

(9)フレーズ単位で索引づけをしたか？：NO

(10)構文解析は行なったか？：NO

(11)シソーラスや用語集などを用いたか？：NO

(12)語義の曖昧性解消は行なったか？：NO

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：NO

(14)誤字脱字やスペルの修正は行なったか？：NO

(15)固有名詞を識別したか？：NO

(16)どのような方法で索引単位に分割したか？：文字列照合

(17) 日本語のヨミを用いたか。NO

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・その他（具体的に）：キーワードとテストコレクションデータとの対応表

(2)索引の概要
・実行ID：KLAB1,KLAB2
・索引の規模［MB］：4.6MB(KLAB1),15MB(KLAB2)
・構築に要した時間［時間］：約24時間(KLAB1),約24時間(KLAB2)
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：YES
・語の出現位置（オフセット）は使用したか？：NO

１．３　NTCIR-1以外の情報源から構築されたデータ:NO

２　検索式の作成

２．１　自動的に作成した検索式:NO

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：2〜3分

(2)誰が検索式を作成したか？
・計算機システムの専門家：（３人）

(3)検索式作成に用いたツール
・知識ベース：与えられた検索課題中の、タイトル、日本語概念部分を利用した。

(4)検索式作成に用いた方法
・ブール演算子(AND, OR, NOT)：ORのみ利用

２．３　対話的な検索式の作成：NO

３　検索

３．１　検索時間

(1)実行ID：KLAB1,KLAB2

(2)検索時間（１検索式に対する平均CPU時間［秒］）：

３．２　検索モデル

(3)その他（具体的に）：概念検索モデル（キーワードの意味に着目して、同義表現（単語）を利用した検索モデル）

３．３　ランクづけの要素：NO

３．４　計算機についての情報

(1)実験に使用した計算機：DELL WORKSTATION WS410MT/K500

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量［GB］：8.5GB

(4)RAMの総容量［MB］：256MB

(5)CPUのクロック数［MHz］：500MHz

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：YES

(以上)

随時検索タスク（Ad hoc IR task）

チーム略称： NTE15

実行ID（複数ある場合はすべて）：NTE151,NTE152

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）： 単語

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語(現実的には n-gram)

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：辞書を用いた極大単語切り出し方法

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：ストップワード以外を選択

(6)語彙（文字）の正規化を行なったか？：YES(ASCIIコードを対応する2byteコードに変換)

(7)ステミングアルゴリズムを用いたか？：NO

(8)語の重みづけを用いたか？：NO

(9)フレーズ単位で索引づけをしたか？：NO
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：NO

(11)シソーラスや用語集などを用いたか？：NO

(12)語義の曖昧性解消は行なったか？：NO

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：NO

(14)誤字脱字やスペルの修正は行なったか？：NO

(15)固有名詞を識別したか？：NO

(16)どのような方法で索引単位に分割したか？：極大単語索引方式

(17) 日本語のヨミを用いたか。NO

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：YES （極大単語索引）
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：NTE151,NTE152
・索引の規模［MB］：317
・構築に要した時間［時間］：0.30
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：自動化
・語の出現位置（オフセット）は使用したか？：YES

１．３　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：(質問の意味がわかりません)
・ファイルの型（シソーラス、知識ベース、辞書など）：辞書
・総記憶量［MB］：33.4
・表現された概念数：約42万
・表現の型：(質問の意味がわかりません)
・構築に要した計算機の稼動時間［時間］：2.5
・構築に要した手作業の時間［時間］：(不明)
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？：YES

(2)外部で構築された補助ファイル（商品含む）：EDR辞書(原辞書として利用)

２　検索式の作成

２．１　自動的に作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：3.585

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語(未知語も含む)
・フレーズの抽出：NO
・構文解析：NO
・語義の曖昧性解消：NO
・固有名詞の識別：NO
・検索式の自動拡張：YES
－シソーラスなど既存のツール：NO
－自動レレバンスフィードバック：NO
＊ローカルコンテクストアナリシス：NO
＊その他（具体的に）：NO
－その他（具体的に）：語の類似度(タームベクトル間の内積)に基づく拡張
・ブール演算子や近接演算子などの自動的付与：YES
・その他（具体的に）：

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：NTE151,NTE152

(2)検索時間（１検索式に対する平均CPU時間［秒］）：NTE151=0.493, NTE152=1.822

３．２　検索モデル

(1)ベクトル空間型を用いたか？：YES

(2)確率型を用いたか？：

(3)その他（具体的に）：

３．３　ランクづけの要素

(1)TF（語の出現頻度）：YES

(2)IDF：YES

(3)その他の重みづけ（具体的に）：文書内共起

(4)意味の近さ：NO

(5)文書中の位置：NO

(6)構文的な手がかり：NO

(7)語の近接（距離）：NO

(8)文書の長さ：YES

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：Sun SS-UE450

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量 [GB] ：90

(4)RAMの総容量 [MB] ：1024

(5)CPUのクロック数 [MHz] ：296

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：YES
・日本語のわかる人がいる：YES
・日本語のわかる人はいない：NO

随時検索タスク（Ad hoc IR task）

チーム略称：R2D2

実行ID（複数ある場合はすべて）：R2D21 R2D22 R2D23 R2D24

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？：単語

(2)英語の索引単位は何か？：単語

(3)日本語文中の英語の索引単位は何か？：単語

(4)どのような方法を用いて索引づけをしたか？：形態素解析

(5)索引語の選択方法は何か？：品詞 ストップワード

(6)語彙（文字）の正規化を行なったか？：YES

(7)ステミングアルゴリズムを用いたか？：NO

(8)語の重みづけを用いたか？：YES

(9)フレーズ単位で索引づけをしたか？：NO

(10)構文解析は行なったか？：NO

(11)シソーラスや用語集などを用いたか？：NO

(12)語義の曖昧性解消は行なったか？：NO

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：NO

(14)誤字脱字やスペルの修正は行なったか？：NO

(15)固有名詞を識別したか？：NO

(16)どのような方法で索引単位に分割したか？：形態素解析

(17) 日本語のヨミを用いたか。：NO

(18) ヨミを用いた場合、ヨミはどのように生成したか。：該当データなし

(19)索引づけに用いたその他の方法（具体的に）：該当データなし

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
- 転置索引：YES (R2D2{13})
- その他（具体的に）：語IDと重みを列挙した単純インデクス
  非転置索引 R2D2{24}

(2)索引の概要
- 実行ID：R2D21
- 索引の規模［MB］：23
- 構築に要した時間［時間］：18
- 実行過程は自動化されているか？：YES
- 語の出現位置（オフセット）は使用したか？：NO

- 実行ID：R2D22
- 索引の規模［MB］：286
- 構築に要した時間［時間］：19
- 実行過程は自動化されているか？：YES
- 語の出現位置（オフセット）は使用したか？：NO

- 実行ID：R2D23
- 索引の規模［MB］：45
- 構築に要した時間［時間］：18
- 実行過程は自動化されているか？：YES
- 語の出現位置（オフセット）は使用したか？：YES

- 実行ID：R2D24
- 索引の規模［MB］：311
- 構築に要した時間［時間］：19
- 実行過程は自動化されているか？：YES
- 語の出現位置（オフセット）は使用したか？：YES

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル：該当データなし

(2)外部で構築された補助ファイル（商品含む）：ChaSen 形態素辞書

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：60

(2)検索式作成に使用した方法
- 索引単位への分割：単語
- フレーズの抽出：NO
- 構文解析：NO
- 語義の曖昧性解消：NO
- 固有名詞の識別：NO
- 検索式の自動拡張：NO
- ブール演算子や近接演算子などの自動的付与：R2D23 R2D24 では近接関係を利用している
- その他（具体的に）：該当データなし

2．2　手動で作成した検索式：該当データなし

2．3　対話的な検索式の作成：該当データなし

3　検索

3．1　検索時間

(1)実行ID：R2D21

(2)検索時間（1検索式に対する平均CPU時間［秒］）：33

(1)実行ID：R2D22

(2)検索時間（1検索式に対する平均CPU時間［秒］）：380

(1)実行ID：R2D23

(2)検索時間（1検索式に対する平均CPU時間［秒］）：53

(1)実行ID：R2D24

(2)検索時間（1検索式に対する平均CPU時間［秒］）：160

3．2　検索モデル

(1)ベクトル空間型を用いたか？：YES

(2)確率型を用いたか？：NO

(3)その他（具体的に）：該当データなし

3．3　ランクづけの要素

(1)TF：YES

(2)IDF：YES

(3)その他の重みづけ：該当データなし

(4)意味の近さ：NO

(5)文書中の位置：NO

(6)構文的な手がかり：NO

(7)語の近接（距離）：R2D23 と R2D24 で使用

(8)文書の長さ：NO

(9)その他（具体的に）：著者キーワードから抽出した文書関連性に基づく文書ベクトル拡張

３．４　計算機についての情報

(1)実験に使用した計算機：Sun ULTRA 2

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量 [GB] ：30

(4)RAMの総容量 [MB] ：128

(5)CPUのクロック数 [MHz] ：296

３．５　その他

(1)上の質問で回答していないシステムの特色：該当データなし

(2)その他（具体的に）：該当データなし

(3)チームの構成員に：
・日本語のnative speakerがいる

Ad hoc IR task

Group's ID:
RMIT

List of Run ID(s):
rmit02

* NTCIR-1 = NACSIS Test Collection 1

1 Indexing

1.1 Indexing

1) Index units for Japanese text (uni-gram, bi-gram, other n-gram, word, phrase, other):

unigram

2) Index units for English text (n-gram, word, phrase, other):

3) Index units for English terms within sentence in Japanese (n-gram, word, phrase, other):

words

4) Method(s) used in indexing (lexicon, morphological analysis, other):

morpological

5) Method(s) used in selection of index terms (e.g.: stop word, type of characters, part of speech, etc.):

all words

6) Standardizing terms (characters)?:

no

7) Stemming Algorithm?:

no

8) Term Weighting?:

yes, standard tf.idf

9) Phrase identification?:
- Kind of phrase:
- Method used (statistical, syntactic, other):

no

10) Syntactic Parsing?:

no

11) Thesaurus and/or lexical resources?:

no

12) Word sense disambiguation?:

no

13) Spelling checking (including manual checking)?:

no

14) Correcting them?:

no

15) Proper noun identification?:

no

16) Method(s) used in tokenizing?:

individual characters

17) Use "YOMI" of Japanese text?:

no

18) (if 17 is Yes) method(s) used to generate "YOMI":

19) Other method(s) used in indexing (Please specify):

1.2 Index data structures built from NTCIR-1

1) Kind of index structures
- Inverted index:
- Clusters:
- Signature files:
- Pat-tree:
- Knowledge bases:
- Other(s) (Please specify):

inverted index

2) Summary of index
- Run ID: rmit02
- Total storage [in MB]: 381
- Total time to build [in hours]: 1
- Automatic process? (If not, number of manual hours [in hours]):
- Use of positional information (off-set)?: no

1.3 Data built from sources other than NTCIR-1

none

1) Internally-built auxiliary files
- Domain:
- Type of file (thesaurus, knowledge base, lexicon, etc.):
- Total storage [in MB]:
- Number of concepts represented:
- Type of representation:
- Total computer time to build [in hours]:
- Total manual time to build [in hours]:
- Total manual time to modify NTCIR-1 (if already built) [in hours]:
- Use of manual labor?:

2) Externally-built auxiliary files (including commodities):

2 Query construction

2.1 Automatically constructed queries

1) Average computer time to build query [in CPU seconds]:
< 1 second

2) Method(s) used in building queries
- Tokenizing (uni-gram, bi-gram, n-gram, word, phrase, other): unigram
- Phrase extraction from topics?: no
- Syntactic parsing?: no
- Word sense disambiguation?: no
- Proper noun identification?: no
- Automatic expansion of queries?: no
  * Preciously-constructed tools such as thesaurus?: no
  * Automatic relevance feedback?: no
    + Local context analysis: no
    + Other(s) (Please specify):
  * Other(s) (Please specify):
- Automatic addition of Boolean/procimity operators?: no
- Other(s) (Please specify):

2.2 Manually constructed queries

NA

1) Average time to construct query [in minutes]:

2) Person constructing queries
- Domain expert:
- Computer system expert:
- Other (Please specify):

3) Tools used in building query
- Word frequency list?:
- Knowledge base?:
- Other lexical tools (e.g. thesaurus, lexicon)? (Please specify):

4) Method used in query construction
- Term weighting?:
- Boolean operators (AND, OR, NOT)?:
- Proximity operators?:
- Addition of terms not included in topics?:
  * Source of additional terms:
- Other (Please specify):

2.3 Interactive queries

NA

1) Initial query constructed automatically or manually:

2) Person doing interaction
- Domain expert:
- Computer system expert:
- Other (Please specify):

3) Average time to complete an interaction [in minutes]):

4) What determines the end of an interaction:

5) Method(s) used in interaction:
- Automatic term reweighting from relevant documents (relevance feedback)?:
- Query expansion from relevant documents (relevance feedback)?:
  * All terms in relevant documents added:
  * Only top X terms added (specify X):
  * User selected terms added:
- Query modification in manual?:

3 Searching

3.1 Search times

1) Run ID: rmit02

2) Computer time to search [average per query, in CPU seconds]:

10 secs

3.2 Searching methods

vector space

1) Vector space model?:

2) Probabilistic model?:

3) Other (Please specify):

3.3 Factors in ranking

tf.idf

1) TF (Term Frequency)?:

2) IDF (Inverse document frequency)?:

3) Other term weights? (Please specify):

4) Semantic closeness?:

5) Positional information in the document?:

6) Syntactic clues?

7) Proximity of terms?:

8) Document length?:

9) Other (Please specify):

3.4 Machine information

1) Machine type for the experiment: PC Pentuim II 333Mhz

2) Was the machine dedicated or shared?: dedicated

3) Amount of hard disk storage [in MB]: 35 000

4) Amount of RAM [in MB]: 256

5) Clock rate of CPU [in MHz]: 333

3.5 Others

1) Brief description of features of your system not answered by above questions:

2) Others (Please specify):

3)Your group has:
- No member who can understand Japanese language:

随時検索タスク（Ad hoc IR task）

チーム略称：STIX

実行ID（複数ある場合はすべて）：STIX1,STIX2

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：品詞

(6)語彙（文字）の正規化を行なったか？：Yes

(7)ステミングアルゴリズムを用いたか？：No

(8)語の重みづけを用いたか？：Yes

(9)フレーズ単位で索引づけをしたか？：No
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：No

(11)シソーラスや用語集などを用いたか？：No

(12)語義の曖昧性解消は行なったか？：No

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：No

(14)誤字脱字やスペルの修正は行なったか？：No

(15)固有名詞を識別したか？：Yes

(16)どのような方法で索引単位に分割したか？：形態素解析と分割用辞書

(17) 日本語のヨミを用いたか。No

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：該当データなし

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：○
・クラスタ：
・シグネチャファイル：○
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：STIX 1 ,STIX2
・索引の規模［MB］：1087,1087
・構築に要した時間［時間］：10,10
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：Yes,Yes
・語の出現位置（オフセット）は使用したか？：Yes,Yes

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：DATABASE
・ファイルの型（シソーラス、知識ベース、辞書など）：関係語辞書
・総記憶量［MB］：0.006
・表現された概念数：335
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：0
・構築に要した手作業の時間［時間］：未計測
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：0
・手作業を行なったか？：該当データなし

・ドメイン：DATABASE
・ファイルの型（シソーラス、知識ベース、辞書など）：タイトルパタン辞書
・総記憶量［MB］：0.003
・表現された概念数：233
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：0
・構築に要した手作業の時間［時間］：未計測
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：0
・手作業を行なったか？：該当データなし

・ドメイン：DATABASE
・ファイルの型（シソーラス、知識ベース、辞書など）：表記揺れ辞書
・総記憶量［MB］：0.01
・表現された概念数：128
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：0
・構築に要した手作業の時間［時間］：未計測
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：0
・手作業を行なったか？：該当データなし

・ドメイン：DATABASE
・ファイルの型（シソーラス、知識ベース、辞書など）：一般語辞書
・総記憶量［MB］：0.002
・表現された概念数：55
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：0
・構築に要した手作業の時間［時間］：未計測
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：0
・手作業を行なったか？：該当データなし

・ドメイン：DATABASE
・ファイルの型（シソーラス、知識ベース、辞書など）：単語bigram辞書
・総記憶量［MB］：1.6
・表現された概念数：814
・表現の型：該当データなし
・構築に要した計算機の稼動時間［時間］：未計測
・構築に要した手作業の時間［時間］：0
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：0
・手作業を行なったか？：該当データなし

(2)外部で構築された補助ファイル（商品含む）：該当データなし

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：未計測

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：その他
・フレーズの抽出：○
・構文解析：×

・語義の曖昧性解消：×
・固有名詞の識別：○
・検索式の自動拡張：×
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：×
・その他（具体的に）：

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：0.5

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：○
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：該当データなし

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：文末処理による体言止め化

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：STIX1,STIX2

(2)検索時間（１検索式に対するCPU時間［秒］）：510,未計測

３．２　検索モデル

(1)ベクトル空間型を用いたか？：Yes,Yes

(2)確率型を用いたか？：No,No

(3)その他（具体的に）：係受けの構造を利用,該当データなし

３．３　ランクづけの要素

(1)TF（語の出現頻度）：No,No

(2)IDF：Yes,Yes

(3)その他の重みづけ（具体的に）：係受けの重要度、類似度,該当データなし

(4)意味の近さ：No,No

(5)文書中の位置：No,No

(6)構文的な手がかり：Yes,No

(7)語の近接（距離）：No,No

(8)文書の長さ：No,No

(9)その他（具体的に）：該当データなし,該当データなし

３．４　計算機についての情報

(1)実験に使用した計算機：SS1000E

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量［GB］：180

(4)RAMの総容量［MB］：1.5

(5)CPUのクロック数［MHz］：85

３．５　その他

(1)上の質問で回答していないシステムの特色：該当データなし

(2)その他（具体的に）：該当データなし

(3)チームの構成員に：
・日本語のnative speakerがいる：Yes
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称： TGL

実行ID（複数ある場合はすべて）： tgl1

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：品詞(名詞)

(6)語彙（文字）の正規化を行なったか？：なし

(7)ステミングアルゴリズムを用いたか？：なし

(8)語の重みづけを用いたか？：あり

(9)フレーズ単位で索引づけをしたか？：していない
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：ない

(11)シソーラスや用語集などを用いたか？：ない

(12)語義の曖昧性解消は行なったか？：ない

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：ない

(14)誤字脱字やスペルの修正は行なったか？：ない

(15)固有名詞を識別したか？：なし

(16)どのような方法で索引単位に分割したか？：文字種分割

(17) 日本語のヨミを用いたか。ない

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：なし

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：転置ファイル(単語のテキスト中の位置情報を含む)
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：

(2)索引の概要
・実行ID：
・索引の規模［MB］：240MB
・構築に要した時間［時間］：12時間
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：自動
・語の出現位置（オフセット）は使用したか？：使用した

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：単語
・フレーズの抽出：
・構文解析：
・語義の曖昧性解消：
・固有名詞の識別：
・検索式の自動拡張：
－シソーラスなど既存のツール：

－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：抽出した語に対するand
・その他（具体的に）：

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：60

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：情報科学部生

(3)検索式作成に用いたツール
・語の出現頻度リスト：なし
・知識ベース：なし
・その他の辞書的ツール（シソーラスや辞書など具体的に）：なし

(4)検索式作成に用いた方法
・語の重みづけ：あり
・ブール演算子(AND, OR, NOT)：あり
・近接演算子：なし
・検索課題に含まれていない語の追加：なし
－追加した語の情報源：
・その他（具体的に）：なし

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：

(2)検索時間（１検索式に対する平均CPU時間［秒］）：

３．２　検索モデル

(1)ベクトル空間型を用いたか？：

(2)確率型を用いたか？：

(3)その他（具体的に）：ブール検索

３．３　ランクづけの要素

(1)TF（語の出現頻度）：検索語の総和

(2)IDF：－

(3)その他の重みづけ（具体的に）：－

(4)意味の近さ：－

(5)文書中の位置：あり

(6)構文的な手がかり：

(7)語の近接（距離）：あり

(8)文書の長さ：

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：Sun Microsysmtems: Ultra2 Model2170

(2)その計算機は専用か共用か：専用

(3)ハードディスクの総容量［GB］：2

(4)RAMの総容量［MB］：128

(5)CPUのクロック数［MHz］：167

３．５　その他

(1)上の質問で回答していないシステムの特色：マルチスレッドによる並列検索

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：いる
・日本語のわかる人がいる：いる
・日本語のわかる人はいない：いる

随時検索タスク（Ad hoc IR task）

チーム略称： UTS

実行ID（複数ある場合はすべて）：UTS1, UTS2

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：
２文字以上の全ての部分文字列（文字n-gram）

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：
２文字以上の全ての部分文字列

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：
２文字以上の全ての部分文字列

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）： Suffix Array

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）： 全ての部分文字列を使用

(6)語彙（文字）の正規化を行なったか？： いいえ

(7)ステミングアルゴリズムを用いたか？： いいえ

(8)語の重みづけを用いたか？： はい

(9)フレーズ単位で索引づけをしたか？： いいえ
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？： いいえ

(11)シソーラスや用語集などを用いたか？： いいえ

(12)語義の曖昧性解消は行なったか？： いいえ

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？： いいえ

(14)誤字脱字やスペルの修正は行なったか？： いいえ

(15)固有名詞を識別したか？： いいえ

(16)どのような方法で索引単位に分割したか？： 全てのSuffix（Suffix Array）

(17) 日本語のヨミを用いたか。 いいえ

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）： Suffix Array

(2)索引の概要
・実行ID： UTS1,UTS2（共通）
・索引の規模［MB］： 1117
・構築に要した時間［時間］： 2
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）： はい
・語の出現位置（オフセット）は使用したか？： はい

１．３　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル：該当データなし
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量［MB］：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）： 該当データなし

２　検索式の作成

２．１　自動的に作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）：
53課題で約1秒

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：
・フレーズの抽出：
・構文解析：
・語義の曖昧性解消：
・固有名詞の識別：
・検索式の自動拡張：
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：

＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：
・その他（具体的に）： 全文

２．２　手動で作成した検索式：該当データなし

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成：該当データなし

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：UTS1
(2)検索時間（１検索式に対する平均CPU時間［秒］）：約100

(1)実行ID：UTS2
(2)検索時間（１検索式に対する平均CPU時間［秒］）：約1000

３．２　検索モデル

(1)ベクトル空間型を用いたか？：はい

(2)確率型を用いたか？：いいえ

(3)その他（具体的に）：文書毎に検索要求の最適な分割を計算する

３．３　ランクづけの要素

(1)TF（語の出現頻度）：使用した

(2)IDF：使用した

(3)その他の重みづけ（具体的に）：部分文字列の長さ

(4)意味の近さ：使用しない

(5)文書中の位置：使用しない

(6)構文的な手がかり：使用しない

(7)語の近接（距離）：使用しない

(8)文書の長さ：使用した

(9)その他（具体的に）：

３．４　計算機についての情報

(1)実験に使用した計算機：SUN Ultra 3000

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量 [GB] ：約100

(4)RAMの総容量 [MB] ：1500

(5)CPUのクロック数 [MHz] ：200

３．５　その他

(1)上の質問で回答していないシステムの特色：

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる： はい
・日本語のわかる人がいる：
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称：WSLab

実行ID（複数ある場合はすべて）：WSLab1, WSLab2

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション 1

1　索引づけ

1．1　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：uni-gram

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：単語

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：形態素解析

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：品詞とストップワード

(6)語彙（文字）の正規化を行なったか？：いいえ

(7)ステミングアルゴリズムを用いたか？：いいえ

(8)語の重みづけを用いたか？：はい

(9)フレーズ単位で索引づけをしたか？：いいえ
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）：

(10)構文解析は行なったか？：いいえ

(11)シソーラスや用語集などを用いたか？：いいえ

(12)語義の曖昧性解消は行なったか？：いいえ

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：いいえ

(14)誤字脱字やスペルの修正は行なったか？：いいえ

(15)固有名詞を識別したか？：いいえ

(16)どのような方法で索引単位に分割したか？：形態素解析

(17) 日本語のヨミを用いたか。：いいえ

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：なし

1．2　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：ベクトル空間法に基づく類似度計算に必要な文書ベクトルと、その転置ファイル

(2)索引の概要
・実行ID：WSLab1, WSLab2
・索引の規模［MB］：213MB, 254MB
・構築に要した時間［時間］：3.5時間, 4.5時間
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）：はい
・語の出現位置（オフセット）は使用したか？：いいえ

1．3　NTCIR-1以外の情報源から構築されたデータ

(1)独自に構築した補助ファイル
・ドメイン：なし
・ファイルの型（シソーラス、知識ベース、辞書など）：なし
・総記憶量［MB］：なし
・表現された概念数：なし
・表現の型：なし
・構築に要した計算機の稼動時間［時間］：なし
・構築に要した手作業の時間［時間］：なし
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：なし
・手作業を行なったか？なし

(2)外部で構築された補助ファイル（商品含む）：なし

2　検索式の作成

2．1　自動的に作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均CPU時間［秒］）：1秒

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：uni-gram
・フレーズの抽出：なし
・構文解析：なし
・語義の曖昧性解消：なし
・固有名詞の識別：なし
・検索式の自動拡張：なし
－シソーラスなど既存のツール：なし
－自動レレバンスフィードバック：なし
＊ローカルコンテクストアナリシス：なし
＊その他（具体的に）：なし
－その他（具体的に）：なし
・ブール演算子や近接演算子などの自動的付与：なし
・その他（具体的に）：なし

2．2　手動で作成した検索式

(1)検索式を作成するのに要した時間（1課題当たりの平均時間［分］）：なし

(2)誰が検索式を作成したか？
・分野の専門家：なし
・計算機システムの専門家：なし
・その他（具体的に）：なし

(3)検索式作成に用いたツール
・語の出現頻度リスト：なし
・知識ベース：なし
・その他の辞書的ツール（シソーラスや辞書など具体的に）：なし

(4)検索式作成に用いた方法
・語の重みづけ：なし
・ブール演算子(AND, OR, NOT)：なし
・近接演算子：なし
・検索課題に含まれていない語の追加：なし
－追加した語の情報源：なし
・その他（具体的に）：なし

2．3　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：なし

(2)誰が検索を実行したか？：なし
・分野の専門家：なし

・計算機システムの専門家：なし
・その他（具体的に）：なし

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：なし

(4)検索を終了した理由は何か？：なし

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：なし
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：なし
－正解文書中のすべての語を追加：なし
－上位Ｘ個の検索語を追加（Ｘはいくつか）：なし
－ユーザが選択した語を追加：なし
・手動での検索式の修正を行なったか？：なし

３　検索

３．１　検索時間

(1)実行ID：WSLab1, WSLab2

(2)検索時間（１検索式に対する平均CPU時間［秒］）：26秒, 48秒

３．２　検索モデル

(1)ベクトル空間型を用いたか？：はい

(2)確率型を用いたか？：いいえ

(3)その他（具体的に）：なし

３．３　ランクづけの要素

(1)TF（語の出現頻度）：はい

(2)IDF：はい

(3)その他の重みづけ（具体的に）：なし

(4)意味の近さ：なし

(5)文書中の位置：なし

(6)構文的な手がかり：なし

(7)語の近接（距離）：なし

(8)文書の長さ：なし

(9)その他（具体的に）：なし

３．４　計算機についての情報

(1)実験に使用した計算機：Sun Ultra Enterprise 450

(2)その計算機は専用か共用か：共用

(3)ハードディスクの総容量 [GB] ：80GB

(4)RAMの総容量 [MB] ：2048MB

(5)CPUのクロック数 [MHz] ：296MHz

３．５　その他

(1)上の質問で回答していないシステムの特色：なし

(2)その他（具体的に）：なし

(3)チームの構成員に：
・日本語のnative speakerがいる：
・日本語のわかる人がいる：はい
・日本語のわかる人はいない：

随時検索タスク（Ad hoc IR task）

チーム略称：
sstut

実行ID（複数ある場合はすべて）：
sstut3

※［ ］内は単位　※ NTCIR-1 ＝ NACSISテストコレクション１

１　索引づけ

１．１　索引づけに用いた方法

(1)日本語の索引単位は何か？（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：n-gramフレーズ.

(2)英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：n-gram.

(3)日本語文中の英語の索引単位は何か？（n-gram, 単語、フレーズ、その他）：

(4)どのような方法を用いて索引づけをしたか？（辞書、形態素解析、その他）：情報量の高い部分.

(5)索引語の選択方法は何か？（例：ストップワード、字種、品詞など）：情報量(idf)

(6)語彙（文字）の正規化を行なったか？：行なわない.

(7)ステミングアルゴリズムを用いたか？：行なわない.

(8)語の重みづけを用いたか？：idfを使用した.

(9)フレーズ単位で索引づけをしたか？：
・フレーズの種類は？：
・フレーズの見つけ方は？（統計的、構文的、その他）:ダイナミックプログラミングによるMaximum scoring pathで決定.

(10)構文解析は行なったか？：行なわない.

(11)シソーラスや用語集などを用いたか？：使用しない.

(12)語義の曖昧性解消は行なったか？：行なわない.

(13)誤字脱字やスペルのチェック（手動も含む）は行なったか？：行なわない.

(14)誤字脱字やスペルの修正は行なったか？：行なわないが，修正と等価な効果のある手法を使用した.

(15)固有名詞を識別したか？：しない.

(16)どのような方法で索引単位に分割したか？：分割しないで部分文字列全体を使用した.

(17) 日本語のヨミを用いたか。:用いない.

(18) ヨミを用いた場合、ヨミはどのように生成したか。

(19)索引づけに用いたその他の方法（具体的に）：

１．２　NTCIR-1から構築された索引データの構造

(1)索引の構造の種類は何か？
・転置索引：
・クラスタ：
・シグネチャファイル：
・Pat木：
・知識ベース：
・その他（具体的に）：単純文字列のMatching.

(2)索引の概要
・実行ID：sstut3
・索引の規模 0[MB] ：
・構築に要した時間0［時間］：
・実行過程は自動化されているか？（自動化されていない場合には、手動での時間数［時間］）0：
・語の出現位置（オフセット）は使用したか？：使用しない.

１．３　NTCIR-1以外の情報源から構築されたデータ:なし.

(1)独自に構築した補助ファイル
・ドメイン：
・ファイルの型（シソーラス、知識ベース、辞書など）：
・総記憶量 [MB] ：
・表現された概念数：
・表現の型：
・構築に要した計算機の稼動時間［時間］：
・構築に要した手作業の時間［時間］：
・NTCIR-1を修正するのに要した手作業の時間（既に構築している場合）：
・手作業を行なったか？

(2)外部で構築された補助ファイル（商品含む）：

２　検索式の作成

２．１　自動的に作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均CPU時間［秒］）0：

(2)検索式作成に使用した方法
・索引単位への分割（uni-gram, bi-gram, その他のn-gram, 単語、フレーズ、その他）：
・フレーズの抽出：
・構文解析：
・語義の曖昧性解消：
・固有名詞の識別：
・検索式の自動拡張：
－シソーラスなど既存のツール：
－自動レレバンスフィードバック：
＊ローカルコンテクストアナリシス：
＊その他（具体的に）：
－その他（具体的に）：
・ブール演算子や近接演算子などの自動的付与：
・その他（具体的に）：なにも行なわない.

２．２　手動で作成した検索式

(1)検索式を作成するのに要した時間（１課題当たりの平均時間［分］）：

(2)誰が検索式を作成したか？
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索式作成に用いたツール
・語の出現頻度リスト：
・知識ベース：
・その他の辞書的ツール（シソーラスや辞書など具体的に）：

(4)検索式作成に用いた方法
・語の重みづけ：
・ブール演算子(AND, OR, NOT)：
・近接演算子：
・検索課題に含まれていない語の追加：
－追加した語の情報源：
・その他（具体的に）：

２．３　対話的な検索式の作成

(1)最初の検索式の作成は自動的か手動か：

(2)誰が検索を実行したか？：
・分野の専門家：
・計算機システムの専門家：
・その他（具体的に）：

(3)検索を完了するまでの時間（１課題当たりの平均時間［分］）：

(4)検索を終了した理由は何か？：

(5)対話で使用される方法
・正解文書からの語の再重みづけを行なったか？（レレバンスフィードバック）：
・正解文書からの検索式の展開を行なったか？（レレバンスフィードバック）：
－正解文書中のすべての語を追加：
－上位Ｘ個の検索語を追加（Ｘはいくつか）：
－ユーザが選択した語を追加：
・手動での検索式の修正を行なったか？：

３　検索

３．１　検索時間

(1)実行ID：sstut3

(2)検索時間（１検索式に対する平均CPU時間［秒］）：約4時間(約14,400秒)

３．２　検索モデル

(1)ベクトル空間型を用いたか？：

(2)確率型を用いたか？：

(3)その他（具体的に）：dpマッチングによるあいまい文字検索.

３．３　ランクづけの要素

(1)TF（語の出現頻度）：使用した．ただし全部分文字列に対して計算する.

(2)IDF：使用した.

(3)その他の重みづけ（具体的に）：なし.

(4)意味の近さ：使用せず.

(5)文書中の位置：使用せず.

(6)構文的な手がかり：使用せず.

(7)語の近接（距離）：使用せず.

(8)文書の長さ：使用せず.

(9)その他（具体的に）：dpマッチングによる類似度を用いた.

３．４　計算機についての情報

(1)実験に使用した計算機：Panasonic CFS21

(2)その計算機は専用か共用か：専用.

(3)ハードディスクの総容量［GB］：3[GB]

(4)RAMの総容量［MB］：96[MB]

(5)CPUのクロック数［MHz］：200MHz/MMX

３．５　その他

(1)上の質問で回答していないシステムの特色：dpマッチングを使用した方法で，一つのベースラインを示す目的で参加した．言語知識を使用しない方法である.

(2)その他（具体的に）：

(3)チームの構成員に：
・日本語のnative speakerがいる：