

# 知的財産利用支援システム研究（客員部門） （岩山真研究室）

Email: iwayama@pi.titech.ac.jp

## （研究分野）

特許情報データベースを効率良く利用することを目的とした情報検索手法、情報構造化手法、情報提示手法について研究を行う。

## （研究テーマ）

### 1) 多様な検索手法

近年知的財産権の重要性はますます高まりつつあり、長年にわたって蓄積されてきた特許情報は、新技術創出、製品開発および経営戦略に欠かすことができないものとなってきている。ところが、この膨大な特許情報データベースから所望の情報を高精度にかつ漏れなく検索することは困難である。現状では単純なキーワード検索に頼っているが、次世代検索技術として様々な検索手法が提案され始めている。本テーマでは、キーワードではなく文書群から直接特許を検索する手法（連想検索）（図1）、検索結果を自動分類して表示する手法（文書クラスタリング）（図2）、検索要求に関連した部分をピンポイントで提示する手法（パッセージ検索）（図3）などを研究する。更には、これらの多様な検索手法を用いつつ、特許文献間、または特許と他文献（例えば論文）間の関連性を自動的に同定し、特許文献集合の相関図を自動的に作成する技術についても研究する（図4、図5）。

### 2) 抄録（要約）の自動作成

特許文献は長いものになると数10ページにも及ぶ。このような特許文献を隈なく読むことは難しいため、通常は出願者または第三者が作成した抄録を読むことが多い。本テーマでは、抄録を自動的に作成する方法について研究する。従来の抄録作成の多くが万人に共通の抄録を作成していたのに対し、本手法では、特許文献検索の利用者に応じて、作成する要約を動的に変化させる技術を目指す（図3）。

### 3) 特許文献の簡単化、構造化

特許文献は独自の書法にのっとっているため、必ずしも読みやすいものではない。特に、「請求項」は一文に全ての情報が詰めこまれるため、重要な項目であるにもかかわらず読みにくい（図6）。本テーマでは、特許文献の可読性を向上させることを目的に、特許文献の文章をより簡素なものに変換する技術を研究する。また、同じ特許文献内で、請求項を簡素な文で説明した箇所も多く、これらの対応付けを自動的に行うことで難読箇所の可読性を高める手法も試みる。

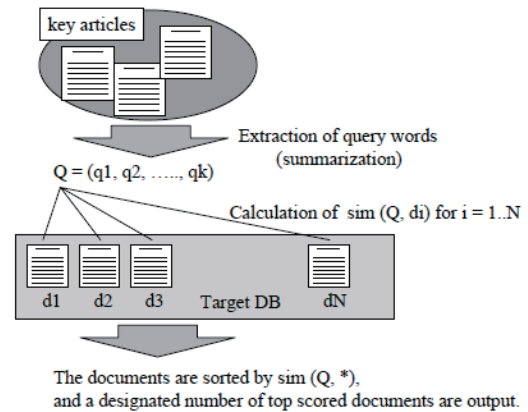


図1 連想検索の概要

Fig.1 Overview of the Associative Search

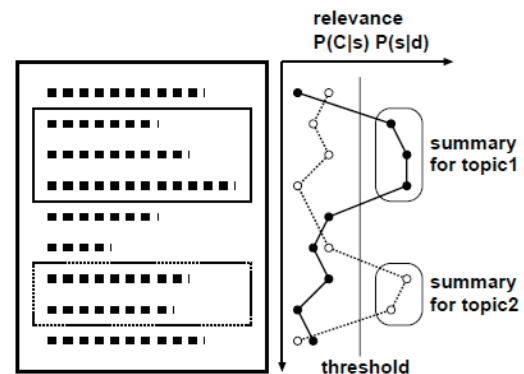


図3 パッセージ検索および抜粋

Fig.3 Passages as Extracts



図5 検索インターフェース

Fig.5 Search Interface

# Intellectual Property Utilization System Division (Guest Chair)

## (Makoto Iwayama Group)

**(Research Field)**

Research on patent retrieval, organization and visualization for the efficient and effective use of patent information.

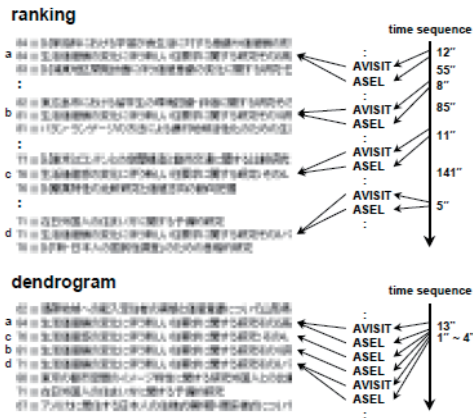


図2 検索結果自動分類表示の効果

Fig.2 The Effect of Automatic Classification of

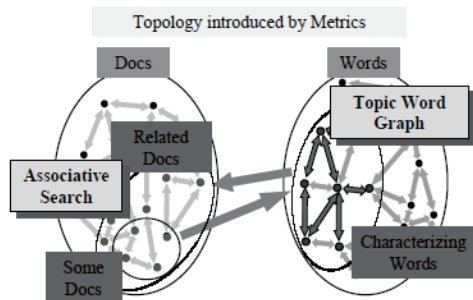


図4 文書間と単語間の関連付け

Fig.4 Association between documents/words

**【特許請求の範囲】**

【請求項1】データベース検索の鍵となる、利用者から入力された文字列形式の質問に対して、前記質問の文字列に類似する展開文字列を表現した展開結果を、予め用意した類似基準に従って求めるに当たり、類似基準は、1ないし2文字の文字列についてその文字列と称号の上で類似関係にある文字を記録した1種以上の類似文字表と、前記展開文字列の各位置の文字が前記質問の文字列から派生する仕方を示す記号である派生型の列のパターンを表す1個以上の正規表現との組で構成され、この類似基準に基づき、前記質問に対する展開文字列の各文字位置について、当該文字の派生に用いた1つ以上の派生型の並んだ派生型リストと前記派生型リストを持つ派生文字リストとの組である派生要素の集合、およびオートマトンの元になる正規表現の集合を求め、次いで、前記正規表現の集合に基づいて、展開文字列の各文字位置で、前記派生要素の許される組み合わせを表す派生型を遷移条件とする有限状態オートマトンを求め、前記派生要素の集合と有限状態オートマトンの組を、前記質問の展開結果として出力することを特徴とする類似文字列展開方法  
(日本国特許2927706「類似文字列の展開方法及びそれらの装置」より)

図6 長い請求項の例

Fig.6 Example of a long claim

**(Current Topics)**

**1) Advanced Patent Retrieval**

Recently the number of submitted patents is extremely increasing and these patents are inevitable for the innovation of new technology, developing new products and the marketing strategy. However, due to the large scale of the patent database, it is difficult for users to find the information they need within the database. We explore new methods for utilizing patent information: searching patents from documents rather than keywords (associative search) (Fig.1), automatic classification of retrieved patents (document clustering) (Fig.2), identifying the relevant part of a patent to the query (passage retrieval) (Fig.3). Integrating these methods, we also try to find the relation between patents or between a patent and research papers, to construct a patent map (Figs. 4 and 5).

**2) Automatical Patent Summarization**

Some patent documents being quite long, for example over dozens of pages, users have the difficulty of reading the whole part of such a long patent. In these cases, users read the extracts written by the authors or other editors. This research aims to automatically create extracts from the original patents. Unlike previous methods which tried to create the single general extract of a document, we try to create usercustomized extracts which depend on users' information needs (Fig.3).

**3) Improving the Readability of Patent Sentences**

Patents are written based on their own style which is largely different from the style of other texts like technical papers. Especially the “claim” part is difficult to understand because it is written in single sentence even if it has several topics (Fig.6). We try to automatically simplify (re-write) the complicated sentence in a claim to raise the readability. In addition, many claims have their explanations in the succeeding sections and these explanations are written in the plain style. We try to automatically find the correspondence between claims and their explanations to help understanding the claim part.