

第III章 既存の情報検索技術

1. 既存の情報検索ツール

1.1 既存の情報検索ツールについての調査結果

研究者向けに、既存の情報検索ツールがどのようなサービスを実施しているかをヒアリングにより調査した。対象としたツールは、アンケートで選択肢とした6種類（JOIS / J Dream、特許電子図書館（IPDL）、PATOLIS、NRIサイバーパテントデスク、Derwent WPI、STN International）にDocuPatを追加した7種類とした。DocuPatを追加した理由は、同ツールが初の試みとして、論文情報と特許情報の両者を検索対象としていることによる。

調査結果を表 III-1 と表 III-2 で示す。科学技術振興機構（JST）については、論文情報を提供している点で他のツールと異なるため、表 III-1 にまとめ、他のツールは表 III-2 にまとめている。

各社とも、研究者をターゲットとしたサービスを展開しており、ユーザー・ニーズに基づく検索機能、表示機能や操作性を備えて大学等に働きかけをしている。DocuPatは、昨年からオプションとして、科学技術振興機構（JST）の論文情報をシステム内に蓄積し、特許情報と共に検索可能としているが、同時に検索は不可となっている。なお、同ツールは今のところ、企業単位の発注によるプライベート・サーバ・サービスとなっており、ID単位での利用はできない。

表 III-1 既存情報検索ツール調査結果(科学技術振興機構(JST))

1. サービスの概要について

項目	内容
(1) 検索オンラインサービスの概要	<p>1. サービス名: JDream、JOIS(2006年4月以降 JDream) (有料)</p> <ul style="list-style-type: none"> 日本最大級の科学技術文献データベースのサービス 収録記事数3,800万件 企業、大学等の検索専門家、研究者・技術者等の他に、病院等医療機関従事者等の方々に利用されている <p>2. STN(有料)</p> <ul style="list-style-type: none"> 日本・米国・ドイツを結ぶ科学技術情報のネットワークシステムで、文献・特許・化学構造・化学反応・数値・全文データベースサービス(CAS系ファイル、DWPIなどを含む) <p>3. J-STORE(無料)</p> <ul style="list-style-type: none"> 大学・公的研究機関、科学技術振興機構(JST)等の技術移転可能な特許情報(出願から1年半未満の未公開特許情報を含む)等を掲載
(2) 科学技術文献情報に関するデータの蓄積範囲	<ul style="list-style-type: none"> 主要サービスの JDream、JOIS 中の代表的なファイル: JSTplusファイルの蓄積範囲 <p>抄録: 日本語及び英語で蓄積、1975年以降現在まで</p> <p>書誌事項: 1975年以降現在まで</p>
(3) 科学技術文献情報サービス以外のサービス	<p>日化辞WEB(無料)</p> <ul style="list-style-type: none"> 有機化合物辞書データベースの検索サービス 名称検索、構造検索が可能 収録物質数は、約220万物質 <p>ReaD(無料)</p> <ul style="list-style-type: none"> 国内の大学・公的研究機関等に関する機関情報、研究者情報等の検索サイト <p>JREC-IN(研究者人材データベース)(無料)</p> <ul style="list-style-type: none"> 研究に関する職を希望する研究者情報と、産学官の研究に関する求人公募情報サイト
(4) サービスのユーザーからの評価	<p>1. JDream、JOIS</p> <ul style="list-style-type: none"> 検索専門家から研究者等が自ら利用するようになり、より使いやすいシステムを求められている JDream のサービスを平成18年4月より開始
(5) サービスのセールスポイント	<p>1. JDream、JOIS</p> <ul style="list-style-type: none"> 科学技術分野を総合的に網羅している 外国語文献も抄録を日本語で検索・参照できる 独自のシソーラスによる統制語付与によりきめの細かい検索が可能 統制語等が今後の解析・可視化ツールの開発に重要な要素を持つと考えている
(6) 科学技術文献情報のフルテキスト検索は可能か	<p>JDream 及び JOIS は検索エンジンにフルテキスト検索エンジンを採用。(検索対象は書誌、抄録データ等)</p>
(7) フルテキスト検索手法	<p>JDream、とも日立製のNグラム方式を採用している</p>
(8) 概念検索を可能としているか	<p>やっていない</p>
(9) 概念検索の仕組み	<p>やっていない</p>
(10) フルテキスト検索および概念検索サービスのユーザーからの評価	<p>1. JDream、JOIS</p> <ul style="list-style-type: none"> JOIS は独自の日本語切り出しツールを用いて検索する機能を有していたが、ヒット件数に問題があった。JDream によるフルテキスト検索エンジン採用でヒット件数が増加(一方ノイズの問題も発生するが、JDream では、各種支援機能によりノイズを減少させる機能を搭載)
(11) 利用しているユーザー数	<p>1. JDream、JOIS</p> <ul style="list-style-type: none"> 発行ID、IPアドレス認証数は、11,000以上 1IDに多数のお客様が共用しており、実際のユーザー数は把握できていない例、1大学のIPアドレス認証等では、学生まで利用が可能

項目	内容
(12) サービスの利用方法	1. JDream, JOIS <ul style="list-style-type: none"> • 申込によりログインID、パスワードを送付 • IPアドレス認証も用意
(13) サービスの英語版	JICST - E (STN) <ul style="list-style-type: none"> • 日本国内文献のみを英語化してサービス
(14) サービスは有料/無料	JDream, JOIS, STN 等は有料 J - STORE、日化辞WEB等は無料

2. 研究者向けサービスについて

項目	内容
(21) 研究者向け検索サービス	基本的に全てのサービスが研究者向けの位置づけ
(22) 研究者向けサービスのデータ蓄積範囲	同じデータベースを利用
(23) 研究者向けと一般向けのサービスの違い	1. JDream, JOIS <ul style="list-style-type: none"> • サービス内容や機能の差は、ない。JDream では、料金体系としては従量料金制度、固定料金制度などのバリエーションを用意している。また、一般者向けの個人用の低額サービス(JDreamPetit)もあるが、機能に制限がある
(24) 研究者向けサービスはフルテキスト検索、概念検索のサービス機能があるか	JDream 及び JDream 等はフルテキスト検索エンジンを採用している
(25) 研究者の評価	データベースは検索のために特殊な技術が必要であったため、研究者は敬遠しがちであった。平成18年4月からサービスする JDream は、検索の特殊な技術を習熟しなくとも検索・分析できる機能を搭載した
(26) 研究者向けサービスにおいて多く利用される検索式(ターム)	JDream では、統制語に誘導する機能や異表記展開辞書、同義語一括検索などの機能を装備
(27) 研究者向けサービスで一番多く利用される出力形式	JDream では、出力項目の見出しを略号から具体的な表記に改めた(例: AU 著者名) ダウンロード形式を増やした(例: CSV) 原文へのリンクを充実した
(28) 研究者向けサービスの利用動向(IP取得数等)	IPアドレス認証による契約が増加していること、学生を含む大学での契約数が伸びていること、企業でも固定料金を採用してきたことにより利用者数は着実に増加している
(29) 利用動向の変動理由	インターネットの普及で、無料サービスが増えたこと、パブル崩壊による企業のリストラに企業図書館や調査担当者が巻き込まれたこと
(30) 研究者向けサービスの重要なポイント	予算管理のし易さ(無料または固定制料金体系)と親しみやすいインターフェイス
(31) 研究者向けサービスを始めた理由	パブル崩壊以降、企業が経営基盤の強化を図り、会社間統合、社内リストラが起り、従来からの情報部門が縮小し、検索専門家が減少した。インターネットの発達とともに、研究者自らが、データベースを検索するようになった

3. サービス全般について

項目	内容
(41) ユーザーからのサービスに対する要望	先行技術調査では、非特許文献も重要であることから文献と特許の統合検索が望まれている
(42) 要望に対する改善予定	特許情報を提供している民間企業等との連携 例: DocuPat文献オプションサービス、科学技術文献アラートサービス
(43) 研究者向けサービスについて、他社のサービスと比較してのセールスポイント	JDream では、研究者向けに分かりやすいインターフェイスと情報検索担当者クラスの検索が初心者でも可能なように設計した
(44) 今後、研究者向けを含めて、サービスの追加、大幅な改善予定	JDream のサービスを予定している

項目	内容
(45) 上記改善の概要	JDream を平成18年4月よりスタートし、今後もバージョンアップを重ねて行く予定。特に解析・可視化ツールの実現等の検討に着手

表 III-2 既存情報検索ツール調査結果(特許文献検索サービス)

提供元	サービスの特徴	非特許文献(学術論文)情報の有無
独立行政法人工業所有権情報・研修館	特許電子図書館(I P D L)は工業所有権情報・研修館が提供する日本唯一の無料特許情報検索サービス。 明治以降の公報全件や外国公報など5,400万件のデータを全て公開している。	なし
株式会社パトリス	国内で最初の知財オンライン検索サービスであるPATOLISを提供している。知財部門や調査会社などの「プロ」が主要な利用者。調査の最後の確認はPATOLISで行うと言われている。独自に公開特許抄録を全件作成したり、用語を統一したキーワードを作成しており、その評価は高い。研究者向けに固定料金制のPATOLIS-Jをサービスし、操作しやすいインターフェイスと質の高い情報を提供している。	審査官が拒絶理由等で引用した非特許文献の蓄積あり[表示のみ]
NRIサイバーパテント株式会社	電子公報を基本としたデータベースにより、当初からインターネットによるサービスを行っている。分かりやすいインターフェイスにより研究者向けに重点を置いている。全文検索、概念検索が主流であり、PATOLISとは違うユーザー層を当初から意識したサービスとなっている。概念検索はNRIが最初にサービスを開始し、その後の流れをリードした。ユーザの要望に対する対処は、サービス開始当初より短期間で行っており、基本的なインターフェイスは変更されていないが、中身は常に向上している。	なし
トムソンコーポレーション株式会社 トムソンサイエンティフィック	世界特許としてのINPADOC・WPIのサービスをはじめ、US・EP・WO・DEの全文を収録した「Delphion」を提供している。 「Delphion」は、特許検索の初心者から上級者までをターゲットとして検索機能を用意し、案件表示画面からリンク機能で特許明細書の内容・関連出願・ファミリー情報・非特許文献を表示できる他、サイテーションツリーをはじめとする統計解析機能も有する。 また、重要な学術論文情報入手できる研究業績評価・研究トレンド分析の「Web of Science」も提供しており、様々なR&Dで使用できる特許・文献データを統合したソリューションとしてのサービスを行っている。 その他多くの特許情報、学術情報ソリューションを提供している。	有り (外国文献のみで、日本語文献は無し)
社団法人化学情報協会	「STN International」は、科学技術分野を中心とした世界中の情報を提供するオンラインサービス。文献、特許情報の他、化学物質の構造・物性・規制情報、医薬品開発情報、ビジネス情報、企業情報など幅広い情報を提供。 検索システムとしては、Web上でコマンドを使わずにメニュー方式で簡単に検索できる「STN Easy」、Web上でコマンドを用いた詳細な検索ができる「STN on the Web」、通信ソフトを使用して、コマンドを用いた詳細な検索ができる「STN Classic」などがある。	有り
日本技術貿易株式会社	富士ゼロックスが開発した「DocuPat」をサービスしている。日本で唯一、特許文献(日本特許と米国特許)と非特許文献(JST科学技術文献)を統合検索UIをベースとした同一サービス内で参照可能なサービス。ただしデータベースは別々であり、検索式の共有は可能であるが、一体的な検索は不可能。 後発システムであり、回答情報の再分析(ワード検索結果からIPC頻度を解析するなど)が可能か点や、研究者に親しみやすいインターフェイスを備えている。	有り JST科学技術文献データ(JST7580,JSTPlus,JMedPlus)合計約2000万件を独自データベースとして搭載。 特許文献と同時検索は不可、特許文献から非特許文献への、また非特許文献から特許文献へのリンクも不可

1. サービスの概要について (1 / 3)

提供元	(1) 特許情報検索オンラインサービスの概要	(2) データの蓄積範囲	(3) 特許情報サービス以外のサービス	(4) 特許情報以外のサービスのユーザーからの評価	(5) サービスのセールスポイント
独立行政法人工業所有権情報・研修館	サービス名: 特許電子図書館 (IPDL) 概要: 公報、経過情報など、基本的に一次情報を無料で提供している	特許・実用新案 公報: 大正 11 年以降、明細書は明治 18 年以降 書誌事項: 平成 2 年以降および平成 10 年以降変化発生分 経過情報: 平成 2 年以降および平成 10 年以降変化発生分	種類: 公開技法 蓄積範囲: 1987 年以降 (最近は社団法人発明協会が許諾した文献のみ)	なし	誰でも 24 時間、無料で必要な情報が得られる
株式会社パトリス	サービス名: PATOLIS-Web・C/W・C/T 概要: 難易度の高い検索や精度の高いデータを求める知財業務従事者に最適なサービス	公報: 公開-昭和 46 年～公告-昭和 30 年～ 独自抄録: 特許公開-昭和 46 年～ 書誌事項: 公開-昭和 46 年～公告-昭和 30 年～ 経過情報: 出願-昭和 39 年～	なし	なし	国内特許・実用新案・意匠・商標および外国データを提供しているオンラインサービス。特許・実用新案については独自の用語統一データを使った技術用語検索、昭和 58 年発行以降の公報全文テキスト検索などが可能。
NRIサイバーパテント株式会社	サービス名: NRIサイバーパテントデスク 概要: 日本で初めてインターネット上で特許情報や技報の検索を可能にしたサービス	公報: 1971 年～ 書誌事項: 1971 年～ 経過情報: 特許庁発行の全データ	種類: 各社の技報 (企業で発行する技術雑誌) 蓄積範囲: 1990 年後半からの分、30 社以上	なし	「簡単」「高速」「情報量」をキーワードに、特許検索の専門家ではない研究者・技術者にも、簡単に利用できるサービスを目指し、日本・米国・PCT (国際公開) 特許情報の共通インターフェイスからの検索・閲覧、審査経過のフロー表示、米国特許の概念検索など、新たな機能・コンテンツの拡充・強化を進めている
トムソンコーポレーション株式会社 トムソンサイエンティフィック	サービス名: Delphion 概要: 知的財産部のみならず、R&Dで利用できる特許・文献データを統合したソリューション	海外情報のみ http://www.thomsonscientific.jp/products/delphion/index.shtml 参照	種類: 学術文献・引用索引データベース『Web of Science』 約 9,000 の高品質な学術雑誌情報と引用文献を収録 蓄積範囲: 自然科学 (SCIE): 1900～ 社会科学 (SSCI): 1956～ 人文学 (AHCI): 1975～	回答なし	一件の出力表示内の各リンクより、関連データを即座に表示できる学術文献へのリンクも可能。
社団法人化学情報協会	サービス名: STN Easy, STN on the web, STN Classic 概要: 特許情報を含むデータベースは約 80 あり、収録国数は 70 ヶ国以上	WPI - 1963 年以降, INPADOC - 1968 年以降, CAplus - 1840 年以降等 データベースによって異なる。	種類: 科学技術全般の文献情報・化学物質情報 蓄積範囲: データベースにより、異なる。	科学技術分野では、世界でもっとも広範に使用されている。歴史もあり、長いおつきあいのユーザーがいる。評価は高いと考える。	全世界から集められた、約 200 のデータベースをご利用いただける。化学物質を中心とした検索機能が特に優れている。
日本技術貿易株式会社	サービス名: DocuPat 概要: 企業単位の契約形態であるプライベートサーバ型で特許/非特許文献の検索・表示が可能なサービス	公報: 日本特許公開年または登録年 1978 年～ 書誌事項: 同上 米国特許: 1980 の全公報 (フルテキスト、明細書、書誌、更新 US クラス) 経過情報: 日本特許出願年 1971 年～	種類: オプションで JST 科学技術文献データ蓄積範囲: 1975 年から現在まで	JST 科学技術文献データはサービスを開始したばかりで具体的な評価はまだない	・豊富なコンテンツとスピーディな公報の閲覧性 ・様々な特許調査の効率化・精度の向上に貢献 ・ブラウザ上で実現される多様な機能 ・特許情報の戦略的な活用を強力に支援 ・国内唯一の特許/非特許文献両対応

1. サービスの概要について(2/3)

提供元	(6) 特許情報のフルテキスト検索は可能か	(7) フルテキスト検索手法	(8) 概念検索は可能か	(9) 概念検索の仕組み	(10) フルテキスト検索および概念検索サービスのユーザーからの評価
独立行政法人工業所有権情報・研修館	不可 ただし、発明の名称、出願人要約、請求の範囲のみで可能	Nグラム方式	なし	なし	公報テキスト検索の利用は多い
株式会社パトリス	可能	非開示	可能(サービス名称は類似文書検索と呼んでいる)	非開示	フルテキスト検索は順調な伸び 類似文書検索は「あたりをつける」程度との評価が伸びていない
NR Iサイバーパテント株式会社	可能。 検索タームには、任意の文字や、文字の組合せが使用できる(半角カタカナを除く) 名詞のほか、「ゆっくり」などの副詞や、「。」や「、」の句読点も使用でき、検索ターム間の論理種(AND条件)や論理和(OR条件)演算も可能	最初はNグラム方式を採用していたが、Nグラムの短所を回避するために、現在は独自方式を採用している	可能 検索項目は「要約」と「特許請求の範囲(実用新案登録請求の範囲)」の2種類 各社発行の技報でも可能	ベクトル演算方式	概念検索は、世の中で最初に始めた 初期の頃は、「何だ? 使い物になるのか?」という反応だったが、最近では定着してきた 先駆者として、信頼は得ていると考えている
トムソンコーポレーション株式会社 トムソンサイエンティフィック	可能	回答なし	不可	なし	回答なし
社団法人化学情報協会	全文特許データベースにおいて可能。 ただし、それ以外のデータベースでもソース文献から専門的な知識を持った情報分析者が抄録や索引の作成を行っているので、効率のよい検索が可能。	キーワード検索	不可	なし	-
日本技術貿易株式会社	可能 「自然文検索機能」ただし、抄録文を対象とした予備検索用のキーワード検索で、該当するIPC候補を選択用に出力する	Nグラムによる全文フルテキストサーチに加えてJST提供のワード切り出しによるキーワード検索機能も利用可能。	可能。日本語自然文による検索機能としてご提供。特許も文献にも自然文検索を同一UIで利用可能としている。 また特許については、ユーザが投入した自然文から候補のIPCが出力され、検索式の作成に利用できる疑似機能有り	ユーザの条件式である日本語自然文から、JST日本語切り出しツールでキーワードをきりだし、あらかじめ同一ツールで切り出しされた公報または文献のキーワードとの適合度をスコアリングして適合度の高い公報または文献から一覧表示する。	分類(IPC)を知らなくても検索ができる点は好評。 またあらゆる書誌情報と自然文との組み合わせが可能である点も好評。 (他社製品では、概念検索と掛け合わせる書誌項目は制限されている) また、特許-文献をあわせた特許文献マップ作成機能をオプションでもっており、評価されている。

1. サービスの概要について (3 / 3)

提供元	(1 1) サービスのユーザー数	(1 2) サービスの利用方法	(1 3) サービスの英語版	(1 4) サービスは有料/無料
独立行政法人工業所有権情報・研修館	不明	インターネット 特許庁または情報・研修館のHPからリンクあり	提供されている ただし、経過情報や公報テキスト検索機能はない 英語版のみのサービスとしてPAJ検索があり、抄録の範囲内ではテキスト検索可能	無料
株式会社パトリス	非公開	ホームページに詳細を掲載	有 (PATOLIS-e)	有料
NRIサイバーパテント株式会社	2300以上	導入説明会を毎月開催 ライターサポートセンターへの問い合わせも可	なし	有料。 基本料金2,000円/月の従量制と、20,000円/月の固定制。また、部門単位、企業単位での利用には、多様な「法人パック」プランあり。
トムソンコーポレーション株式会社 トムソンサイエンティフィック	非開示	ホームページに詳細を掲載	有り (日本語版無し)	有料のビジネスとして展開しており、今回の調査のような動向は、民間としては大変に危惧するところ こちらの情報提供が、IPDLの改善等に繋がるのなら、積極的な協力は避けたい
社団法人化学情報協会	-	ホームページに詳細を掲載。 ヘルプデスク等、問い合わせ窓口あり。	基本は、英語。 日本語や独語、仏語のデータが表示されるデータベースもある。	有料。 料金は、インターフェースやデータベースによって異なる。
日本技術貿易株式会社	20数社	担当営業に連絡。B3	なし	有料 固定料金制で納得いくまで、繰り返し行う事ができる安心感と、原本取り寄せが減る経済的メリットを生む

2. 研究者向けサービスについて (1 / 3)

提供元	(2 1) 研究者向けサービス	(2 2) 研究者向けサービスのデータ蓄積範囲	(2 3) 研究者向けと一般向けのサービスの違い	(2 4) 研究者向けサービスには特許公報のフルテキスト検索、概念検索のサービス機能があるか	(2 5) 研究者の評価
独立行政法人工業所有権情報・研修館	なし	なし	なし	なし	なし
株式会社パトリス	サービス名：PATOLIS-J 概要：研究者・開発者が自身の研究・開発を目的として利用することを前提としたシステム。	同じ	有 操作が簡単、完全固定料金制、マーク機能(選択した文献の保存機能)	ともに有	直接は聞こえない (知財部のフィルタがかかってしまう)
NR Iサイバーパテント株式会社	項番(1)に同じ。 製品コンセプトとターゲットは、研究者。	同じ	なし	有	項番(1 0)に同じ。
トムソンコーポレーション株式会社 トムソンサイエンティフィック	主なサービス名： Delphion、Web of Science 概要：製品コンセプトが研究者向けである その他研究者向け学術情報プラットフォームWeb of Knowledgeを提供	同じ(非特許文献にも対応)	なし ただし、Delphion の検索においては、初級・中級・上級の検索画面を用意しており、研究者には中級を勧めているが、好評	フルテキスト検索有(海外データのみ提供)	回答なし
社団法人化学情報協会	サービス名： SciFinder、SciFinder Scholar 概要：研究者向けである。特許情報・科学技術文献情報を区別せずに検索できる。	同じ。 種類：科学技術文献と特許情報 蓄積範囲：日本と外国を合わせ約50ヶ国	GUI形式で、検索に使用するキーワードや化学構造図の入力を簡素化。	キーワードを入力すると自動的に同義語を認識する。	操作性がよく、評価は高い。
日本技術貿易株式会社	サービス名：DocuPat 概要：製品コンセプトが研究者向けである。	(1)と同様 (非特許文献にも対応)	なし	有	引き合いが数社あり(成立は未) 1.0月から開始したサービスのため、ユーザの詳細動向は未把握(以下、項番(29)まで)

2. 研究者向けサービスについて (2 / 3)

提供元	(26) 研究者向けサービスにおいて多く利用される検索式 (ターム)	(27) 研究者向けサービスで多く利用される出力形式 (公報形式など)	(28) 研究者向けサービスの利用動向 (ID取得数等)	(29) 利用動向の変動理由	(30) 研究者向けサービスの一番重要なポイント
独立行政法人工業所有権情報・研修館	なし	なし	特許電子図書館 (IPDL) 全体としては増え続けている	ヘルプデスクの質問が、開始当初から「オペレーションに関する問い合わせ」が一番多く、現在もその傾向が変わっていないことから、利用者層が増え続けていると考えられる	なし
株式会社パトリス	公報全文テキスト検索 知財調査になじみのない研究者にとっては、思いついた技術用語を全文テキスト検索するのが最もやりやすく、理解もしやすいため	公報形式、出願人抄録形式	微増	知財調査が知財部門の専門業務ではなく、設計、研究、開発などを行う他の部門でも行われるようになってきたため	直にドキュメントにたどりつける 文献のダウンロードが容易
NR I サイバーパテント株式会社	フルテキスト検索、概念検索	一覧表示、抄録表示、全文表示の順	上昇傾向。 メーカーの研究者はもとより、IT系企業や大学のユーザが増えてきた	最近の知財部は、知財戦略や知財活用の方面に忙しく、調査に手が回らない状況にある為、研究者による検索が増えていると考える	・知財部から見れば邪道でも、研究者にとっては良いものがあるという視点 ・直感的に使えるもの。 数ヶ月に一度のトライでも、違和感をもたせない操作性 ・簡素化を追求する事と、見た目に大きな変更を施さない事
トムソンコーポレーション株式会社 トムソンサイエンティフィック	非開示	非開示	非開示	非開示	簡易性 学術文献をサービスしている点からも、特に、特許用語を意識させない工夫が必要と考えている
社団法人化学情報協会	技術的なキーワード。 化学構造質問式。	抄録。 特許と雑誌論文の出力形式を同等に出力、リンクにより、ソースの表示に移行する。	数は不明。 他のサービスと比べると、近年かなりの伸びがある。	操作性の良さが受け入れられたと考えている。	直感的に使用できる点。
日本技術貿易株式会社	研究所・将来技術の技術企画部門では、日本特許に加えて米国特許、文献を相互に参照・利用している。DocuPatについては、日米特許+文献を統合UIで検索が可能であり、日本特許では全文検索キーワード、IPC、Fタームが、米国特許ではキーワードと更新USクラスが中心。また文献ではキーワードが中心となる。	(1) CSV形式により、検索集合を一括保存する機能。エクセルなどで簡単に分析が可能。DocuPatでは日本特許49項目、文献34項目、米国特許63項目と豊富な機能をもっている。 (2) 統合マップ形式：特許と文献に同一UIで独自分類やメモを付与して、統合マップとして出力。	未把握	企業内においては、特許出願が開発の副産物としてなされる時代から、企業の将来の付加価値増大のための戦略出願に重点がおかれつつあり、将来技術のサーベイとして特許と併用する形で文献のニーズが増大してきており、このようなニーズにこたえるために日米特許に加えて文献を統合UIのなかで利用できるシステム(サービス)を商品化した。	国内唯一の特許/非特許両文献対応、検索式の転用による横断的な検索であり、DocuPatでは日米特許と文献の検索を統合UIで相互利用可能。 「大きな流れを見せるが、検索時の手法には深く関わらせない」、「足を踏み入れさせない部分があっても良い」との観点から、検索支援機能に重点を置く

2. 研究者向けサービスについて (3 / 3)

提供元	(3 1) 研究者向けサービスを始めた理由
独立行政法人工業所有権情報・研修館	なし
株式会社パトリス	知財調査が知財部門の専門業務ではなくなり、設計、研究、開発などを行う他の部門でも行われるようになってきたため
N R I サイバーパテント株式会社	サービスを開始した1997年当時は、知財にまじめに取り組んでいるのは、大手メーカーの知財部程度だったが、近い将来には、調査は研究者が直接行う時代が来ると読んでいた
トムソンコーポレーション株式会社 トムソンサイエンティフィック	回答なし
社団法人化学情報協会	各研究室/PC にインストールする事で、情報担当者や図書館でなくとも必要な情報が得られる様な環境作りが必要と考えた。
日本技術貿易株式会社	潜在的ニーズの高まり 公的機関・国立大学の法人化や企業人の取り込み、企業側では社内教育の徹底による調査の徹底に加えて、戦略出願のための将来技術サーベイ等がバックボーンとしてある

3. サービス全般について

提供元	(41) 研究者向けを含めて、ユーザーからのサービスに対する要望	(42) その要望に対する改善予定	(43) 研究者向けサービスのセールスポイント	(44) 研究者向けを含めて、サービスの追加、大幅な改善予定	(45) 改善の概要
独立行政法人工業所有権情報・研修館	要望点：公報PDF出力の文献単位出力、応答速度の向上、蓄積データ範囲の拡大など	未定	なし	一部対応予定	未定
株式会社パトリス	サービス種類：PATOLIS-Web・C/W・C/T、PATOLIS-J 要望点：今後はマニュアルレス操作・コンテンツのわかりやすさ・一般技術文献データの包含などが求められ、課題になると考えられる	PATOLIS-Web：画面および画面遷移を変更して操作性を改善	公報テキストそのものを全文検索することもできるが、審査上の最新情報を検索することもできるPATOLIS-Jには検索結果にマークをつけておく機能がある	未定	未定
NRIサイバーパテント株式会社	要望点：データの蓄積範囲には、常に適及を求められる条件を満たせば、更に先への適及の声がある	自社での作成を含め、あらゆる情報源を対象に検索している	項番(5)と同じ	営業部門にも、知財情報が必要とされているので、その方面も狙い目	抜本的な改造ではなく、日々、様子を見ながら新たなものを提供していく方針
トムソンコーポレーション株式会社 トムソンサイエンティフィック	要望点：非開示	非開示	特許から論文、論文から特許とも、文献の引用・被引用の関係がすぐにビジュアルでわかり、簡単に参照可能で相互利用を推進 独自に長期(数十年)蓄積したデータで実現	あるが、詳細は言えない	言えない
社団法人化学情報協会	サービス種類： 要望点：データや機能面での強化要望がある。	要望は、優先順位の高いものから反映する方針。ECLAやFERMIは、ユーザーの要望により、追加した。	特許と雑誌論文を同時に検索が出来る。	検索結果の分析機能	なし
日本技術貿易株式会社	サービス種類：DocuPat 要望点：非特許文献検索インターフェースの特許検索ライク化 (特許検索で使った検索式そのもので、非特許文献を検索できる様に発展して、双方向に)特許分析、文献のマクロ分析+テキストマイニング技術による分析の要望が最近強くなってきている。	非特許文献を特許の分類で検索できるようにする検討を行っている 特許-文献の分析については、個別対応で対応していく方針。	(30)と同じ	予定しているターゲットは研究者であり、研究者にとって1ツールで全てまかなえるように改善を行ってきている。	検索機能とあわせて特許-文献の分析機能(サイテーション・クラスター分析など)を計画している。検索-公報表示の現状DocuPat2.Xから、R&D情報分析ステーション化を検討しており、個別対応の体制を整えつつある。

1.2 研究者側のニーズと既存の情報検索ツールとの比較

アンケートの結果による研究者側のニーズを、既存の情報検索ツールのサービスがどの程度満足しているか表したものが、表 III-3 である。アンケートの Q. 12 から Q. 15 までを研究者側のニーズとみなし、要望の多かった順に色分けしている。要望度が一番高い項目をピンク、二番目に高い項目をブルー、三番目に高い項目をイエローとしている。各サービスが要望項目を既に満たしている場合は、満たしていない場合は×、一部満たしている場合は とし、要望度の高い項目を満たしている場合は色つきの としている。結果として、色つきの が多いサービスは、現時点で研究者の要望をある程度満たしていると考えられる。特許電子図書館 (IPDL) は特許庁の保有する一次情報を広く民間に提供するという使命のため、必ずしも研究者の要望を満たしているとは言えない。

表 III-3 研究者側のニーズと既存の情報検索ツールとの比較

Q.12 必要な文献/情報(データ)の種類

研究者のニーズ	アンケートによる要望度(%)	特許電子図書館(IPDL) (独立行政法人工業所有権情報・研修館)	PATOLIS-J (株式会社パトリス)	NRIサイバーパテントデスク (NRIサイバーパテント株式会社)	Delphion (トムソンコーポレーション株式会社・トムソンサイエンティフィック)	STN International (社団法人化学情報協会)	DocuPat (日本技術貿易株式会社)
特実、意匠、商標全文献	36.7		×	×	×	×	×
特許文献のみ	29.9						
外国の文献	41.3		×				×
パテントファミリー	29.7	×	×	×			×
リーガルステータス	13.1						
特許文献とリンクした論文等	53.7	×	×	×	(一部あり)		
関連する(引用/被引用)特許文献	25.8	(経過情報のみ)	×				

Q.13 必要な情報(データ)の範囲

研究者のニーズ	アンケートによる要望度(%)	特許電子図書館(IPDL) (独立行政法人工業所有権情報・研修館)	PATOLIS-J (株式会社パトリス)	NRIサイバーパテントデスク (NRIサイバーパテント株式会社)	Delphion (トムソンコーポレーション株式会社・トムソンサイエンティフィック)	STN International (社団法人化学情報協会)	DocuPat (日本技術貿易株式会社)
現存する全文献	39.0		×	×	×	×	×
過去 20 年以内の文献	57.8						

Q.14 必要な情報検索等の機能

研究者のニーズ	アンケートによる要望度(%)	特許電子図書館(IPDL) (独立行政法人工業所有権情報・研修館)	PATOLIS-J (株式会社パトリス)	NRIサイバーパテントデスク (NRIサイバーパテント株式会社)	Delphion (トムソンコーポレーション株式会社・トムソンサイエンティフィック)	STN International (社団法人化学情報協会)	DocuPat (日本技術貿易株式会社)
全文フルテキスト検索機能	59.0	×				(統制語による)	

研究者のニーズ	アンケートによる要望度(%)	特許電子図書館(IPDL) (独立行政法人工業所有権情報・研修館)	PATOLIS-J (株式会社パトリス)	NRI サイバーパテントデスク (NRIサイバーパテント株式会社)	Delphion (トムソンコーポレーション株式会社・トムソンサイエンティフィック)	STN International (社団法人化学情報協会)	DocuPat (日本技術貿易株式会社)
領域を特定したテキスト検索機能	31.2						
類似検索機能(同義語・類義語)	43.7	×	×			×	×
キーワードやシソーラス	50.3	×					
分類検索機能(IPC・Fターム)	5.4						
分類検索支援機能	22.0	(別サービス)	(別サービス)	×	×		
書誌事項検索機能	26.7					(一部あり)	
請求の範囲の構造明示機能	23.6	×	×	×	×	×	×
請求の範囲と詳細な説明リンク機能	22.9	×	×	×	×	×	×

Q.15 必要な操作性

研究者のニーズ	アンケートによる要望度(%)	特許電子図書館(IPDL) (独立行政法人工業所有権情報・研修館)	PATOLIS-J (株式会社パトリス)	NRI サイバーパテントデスク (NRIサイバーパテント株式会社)	Delphion (トムソンコーポレーション株式会社・トムソンサイエンティフィック)	STN International (社団法人化学情報協会)	DocuPat (日本技術貿易株式会社)
ストレスを感じさせない検索速度	75.0	(混雑中あり)					
検索用語ハイライト機能	27.9		×		×		
回答結果、文献の編集機能	25.4	×	×	×			×
回答一括印刷、ダウンロード機能	56.9	(文献単位)					
回答保存機能	26.9	×					
機械翻訳機能	17.3	(日本語英語のみ)	×	×	×	×	×

これら機能については2006年3月末に特許電子図書館(IPDL)において機能改善が行われた。

2 . 現在の検索技術等の状況

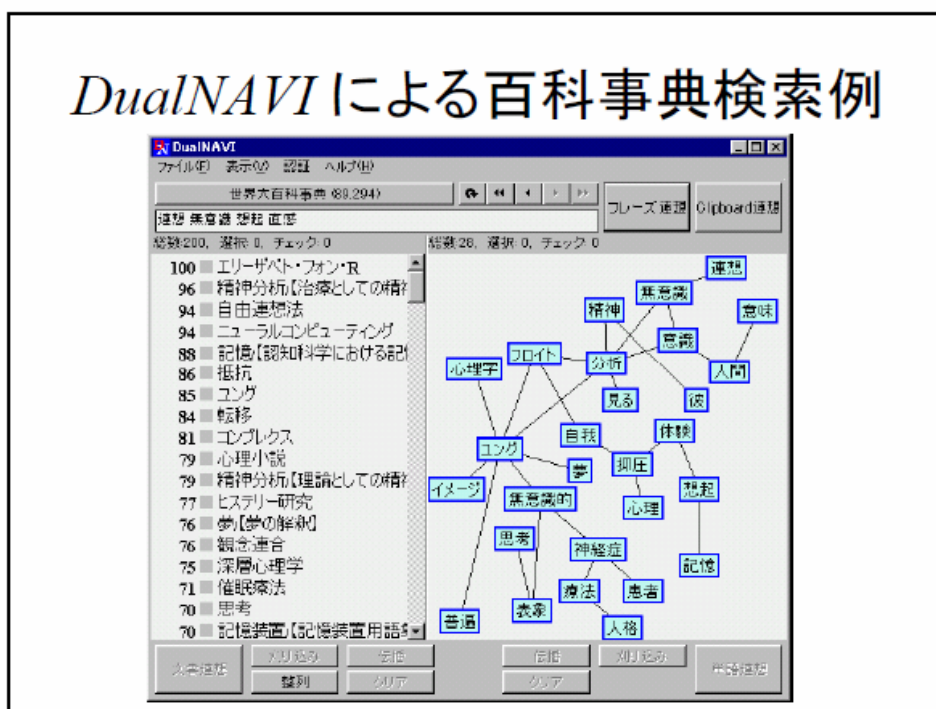
2 . 1 連想する情報サービス

基本コンセプトである「連想の情報学」では2つの「連想」を考える。1つは人間の頭の中で起きる連想である。人は自分が生まれてからの膨大な記憶を持っているが意識的には思い出せない。しかし、何か知的活動するためには関連する事柄を適切に想起することが欠かせない。もう1つは情報空間における連想である。そこにも探しだせないほどの情報があり、その中から適切な情報を集めることで何か面白いことができる訳で、これは人間の頭の中の状況に似ている。この2つの連想を適切に結びつけることができれば、人間の連想する力は高まり、より創造的な仕事が可能になると考えられる。

10年前からこのアプローチで研究されており、DualNAVIという文書検索システムが作られた。DualNAVIの開発経験から、連想の情報学にとってコンピュータ・サイエンス的に一番本質的な計算機構が考えられ、独立のソフトウェアとして切り出されたのがGETAである。GETAは1,000万件を越える大規模データベースを対象に、情報の近さを非常に高速に判定し、関連情報を瞬時に見つけることができる。

DualNAVIは、左側にドキュメントの空間、右側に単語の空間があり、左側のドキュメント群の要約が右側で言葉のグラフとして表示される。人間の記憶は言葉をノードとするグラフ構造に関連づけられていると考えられているが、それと類似の構造をもつこのグラフが人間の頭の中の記憶を呼び起こす。

図 III-1 DualNAVIによる百科事典検索例

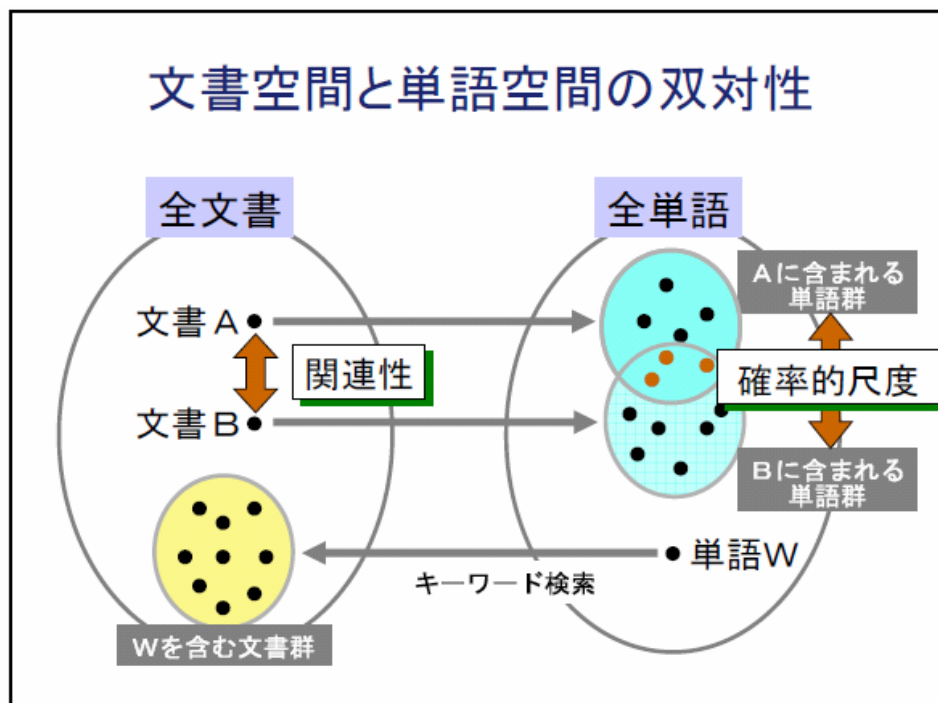


DualNAVIの基礎となった考え方を説明する。左側に全文書、右側に全単語の空間を考える。文書空間は、ひと昔前まではデータベースに綺麗に分かれていたが、Webの出現によりあらゆる種類の文書が混在する状況となった。これらをどうすれば统一的に扱うことができるかを考える必要がある。情報検索の分野では、文書を単語の頻度つきの集まりとして扱うモデル化がよく用いられる。この対応づけの逆を取り、どの単語がどの文書に何回現れたかという情報を集めることにより、単語は頻度つきのドキュメントの集合に対応づけられる。これにより、2つの集合はある意味でコインの裏表のようにして結ばれる。

文書と文書の距離、すなわち関連性は、対応する単語の頻度つき集合間の距離としてモデル化できる。単語空間で定義されるこの距離は、単語の統計的な振る舞いを考慮した数学的に綺麗な枠組みで多くの研究がなされている。これにより、従来の秩序が失われてしまった文書空間に関連性に基礎づけられた秩序が導入される。逆向きの対応づけ

により、単語空間にも距離が定義される。

図 III-2 文書空間と単語空間の双対性

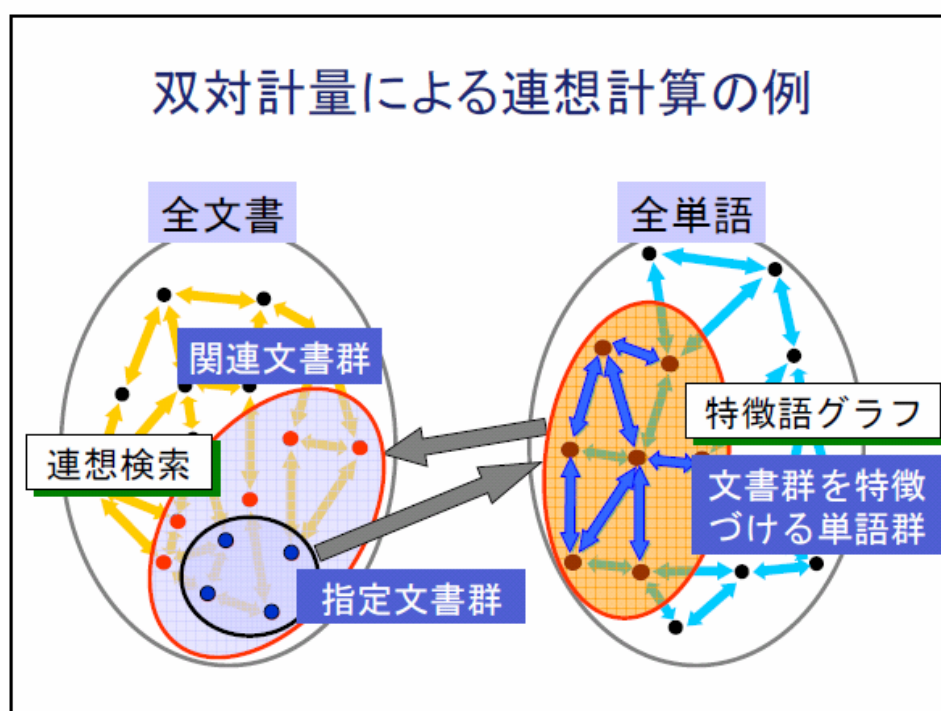


重要なのは、これら 2 つの距離が相互に矛盾しない関係を保っていることである。近い単語は同じドキュメントに含まれやすく、近いドキュメントは同じ単語を含みやすい。

この秩序が連想の情報学に役立つことを示す。例えば、ある 4 つのドキュメントに興味があるとする。それらを単語空間に移すことにより、対応する頻度つき集合を重ね合わせて 1 つの頻度つきの集合が得られる。ここで薄い矢印は元々のベースラインの関連性を、濃い矢印は選択された 4 つのドキュメント中での単語の関連性を表現している。単語と濃い矢印だけを残してグラフを表示すると、文書群の内容を感度よく要約するものになる。これを「特徴語グラフ」と呼び、文書群の自動要約機能と捉えている。特徴語グラフを逆向きの対応で文書集合へ引き戻すことにより、ドキュメントの集合が得られるが、これらは元々の 4 つの文書と内容的に近い文書群と考えられる。これら 2 つの矢印を合成することにより、ドキュメント・イン、ドキュメント・

アウトの検索が実現できる。これを文書連想検索と呼んでいる。このような計算を高速に実行するために開発されたのがGETAである。GETAはIPAから援助を得て3年間で作成され、1,000万件規模のDBに適応可能なソフトウェアとして2002年8月から配布されている。

図 III-3 双対計量による連想計算の例



連想検索のメリットは、検索で使われる文脈の連続性である。例えばある事柄について百科事典と新聞を検索する場合を考える。普通であれば、まず百科事典を検索し、その検索結果を頭に残して、次に新聞DBについてまた初めから検索しなければならない。連想検索では、百科事典を検索した結果を丸ごと質問文として、新聞DBを検索することが可能となる。すでに正解と判定した文書群を質問文として使えることから、検索が進むにつれ、次のデータベースを探す精度が上がるというメリットがある。例えば研究者は論文について非常に土地勘があるので、まず論文DBを検索して自分の調べたいことを捉えておき、それを基点として、特許DBを効率よく検索できると考えられる。

G E T A はすでにいくつかの実用的な情報サービスに使われている。代表的なものとしては、国立情報学研究所が提供する W e b c a t P l u s という図書情報検索システムである。1,000館の大学図書館の収蔵情報をまとめて検索できるサービスとして好評である。連想検索では、新聞記事を丸ごと貼り付けて検索することもできる。すべての自立語が検索に使われるためヒット件数は膨大になるが、検索結果の並び順は、新聞記事と近い順なので、通常は上位数画面を眺めれば十分である。検索結果から特に興味のある本にチェックをつけて再検索することにより、選んだ本と近い本が検索される。対象が特許に変わっても全く同じようにできる。

W e b c a t P l u s の連想検索結果には、上位40冊の内容を要約する単語30個が抽出されて画面右にリストとして表示される。これらのいくつかにチェックをつけて再検索することにより、これらの言葉で検索できる。

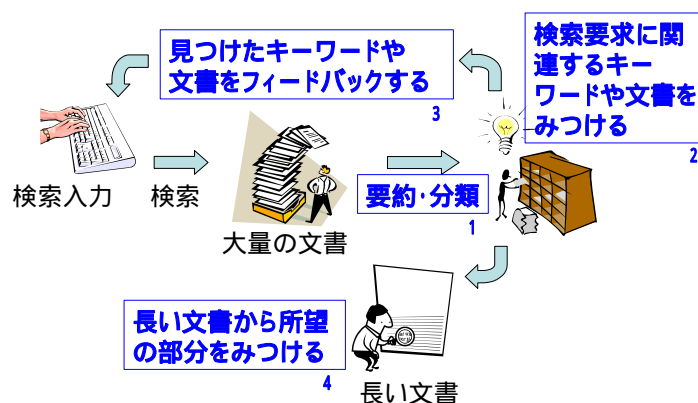
最後に特許の例を述べる。百科事典で「宇宙」を検索すると、いくつかの項目が表示される。その中で興味のある項目を選び、検索対象を特許DBに切り替えて連想検索とすると、百科事典の文章を質問文として特許を検索し、関係しそうな特許が表示される。その中に面白そうな特許が見つければ、チェックして連想することにより、今度は特許同士の近さで検索される。検索対象を請求項や要約の部分や、実施例の部分に絞って行えば、よりシャープに検索できると考えられる。

2.2 特許検索・読解支援インターフェイス

特許検索・読解支援のためのインターフェイス Dual NAVI について研究が進められている。Dual NAVI とは、検索結果を文書空間と単語空間に並置して表示するインターフェイスで、両空間を相互に渡り歩くことで所望の情報を効率よくみつけることができる。また Dual NAVI は、特許のような長い文書の読解を支援するため、文書中のトピック分布を表示するインターフェイスも備えている。

検索の流れを考えると、一回の検索で所望の情報にたどり着くことは少ない。ユーザーは、検索結果から何らかの情報を得て、その情報をシステムにフィードバックしながら徐々に検索結果を精錬していく(図 III-4)。この過程で、システムは図中の 4 つの機能を有する必要がある。1 から 3 までは検索支援に関する機能であり、4 は読解支援に関する機能である。以下ではそれぞれの機能について説明する。

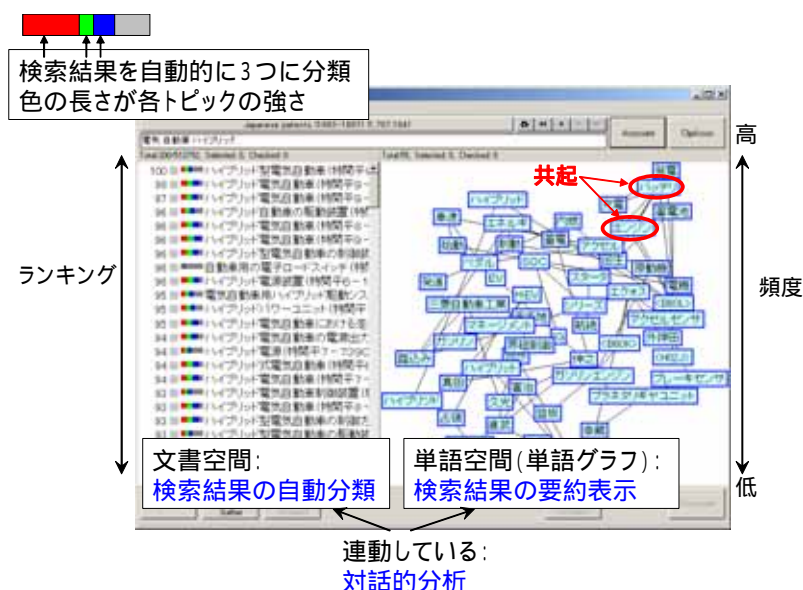
図 III-4 検索・読解の流れ



2.2.1 検索支援

まず、検索結果の概要をユーザーが把握するために、システムは大量の検索結果を要約・分類せねばならない。図 III-5 に Dual NAVI における検索結果表示画面を示す。

図 III-5 DualNAVIの検索結果表示画面



左側が文書空間で、通常は検索入力に近い順に検索結果の文書タイトルが並んでいる。DualNAVIでは、検索結果から3つの主要トピックを自動抽出し、各文書タイトルの左に、その文書が各トピックに属する度合いをカラーバーで表示する。このカラーバーをトピックバーと呼ぶ。各トピックの内容はそれぞれの特徴単語リストを見ることで把握できる。トピックバーを使うことで、ユーザーは検索結果から興味のあるトピックを含む文書のみ注目することができる。

右側の単語空間は、検索結果の文書群に特徴的に表れる単語をグラフ表示している。ここでは、共起している（同じ文書に現れやすい）単語間にリンクが張られている。また、縦軸は単語の出現頻度を表しており、上に位置する単語ほど頻度の高い一般語となる。単語空間は検索結果を要約していることに相当する。単語グラフを調べることで、ユーザーは当初思い浮かばなかったキーワードを見つけることもできる。

文書空間と単語空間は連動している。例えば、文書空間において文書をチェックすると、その文書に含まれている単語が単語グラフにおいてマーキングされる。つまり、検索結果全体の要約である単語グラ

フ中で自分が指定した文書がどこに位置しているのかがわかる。その過程で自分が気付かなかった検索キーワードをグラフ上で見つけた場合、それらを直接検索式にフィードバックして、新たな検索を行うこともできる。

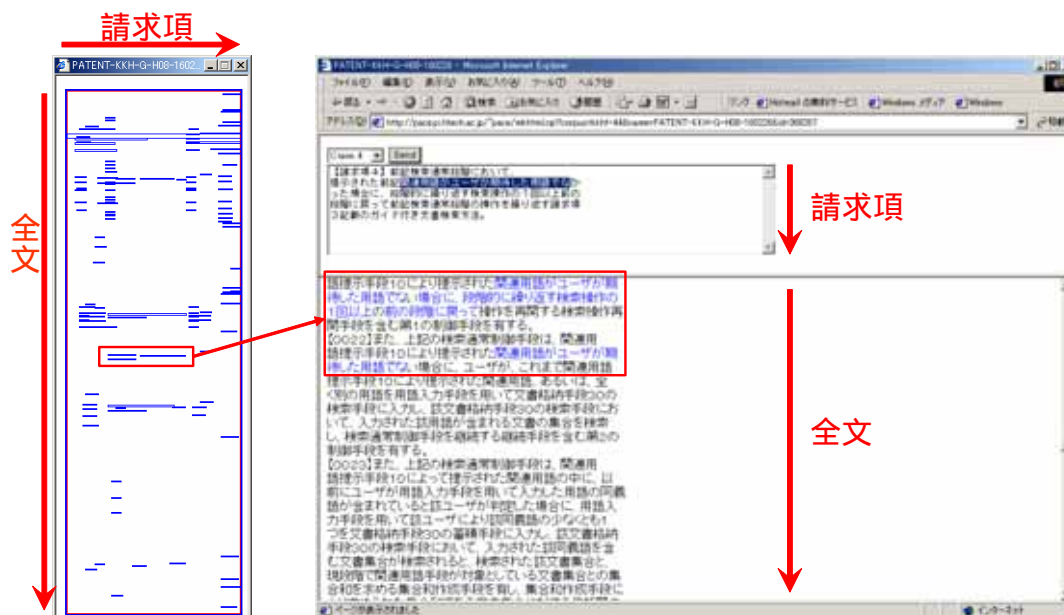
逆に、単語空間において興味のある単語を指定すると、その単語を含む文書の左にチェックマークが付く。チェックマークが太いほど、指定した単語を多く含むことを意味する。よって、チェックマークの太さで文書を並べ替えることで、自分が興味を持った単語を多く含む文書を上位に集めることができる。この過程で正解文書をいくつか見つけた場合は、見つけた文書をそのまま検索式にフィードバックし、見つけた文書に近い文書を集めることができる。

以上のように、DualNAVIでは文書空間と単語空間のインタラクションと、単語・文書からのフィードバックにより、効率良く検索結果を洗練することができる。

2.2.2 読解支援

今回のアンケート調査でも、特許を読解する上では、請求項、課題、効果の部分が重要だという意見が多かった。この中でも特に請求項が重要であるが、請求項のみから発明の内容を把握することは困難であり、請求項に対応する実施例を読まねばならないことが多い。ところが特許は長いため、目的の場所を上から順に探していたのでは時間がかかってしまう。そこで、請求項に対応する部分を自動的に見つけるためのインターフェイスについて研究している。具体的には、明細書全文中に請求項の文字列がどのように分布しているのかをトピックマップという二次元マップで図示する。図 111-6 にトピックマップの例を示す。

図 III-6 トピックマップ



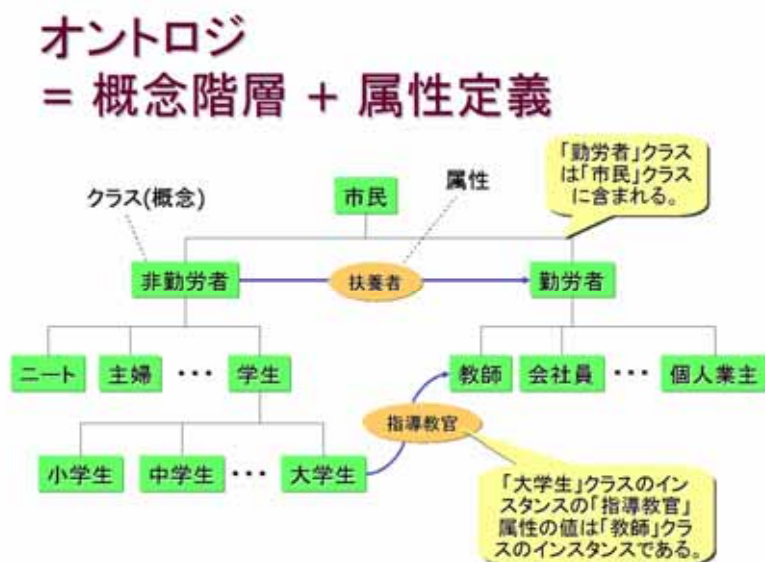
トピックマップでは、請求項が横軸に、特許全文が縦軸に相当する。左端が請求項の最初の文字、右端が最後の文字、上端が特許全文の最初の文字、下端が最後の文字になる。トピックマップ中の矩形（実際には縮尺の関係で線に見える）は、各矩形の辺に相当する請求項、特許全文の部分が対応していることを表している。トピックマップを見ることで、請求項のトピックが特許全文中にどのように分布しているのかを一目で把握することができる。注目した対応部分をクリックすれば、その対応箇所を直接頭出しすることもできる。これは、請求項による目次に相当する。トピックマップを使うことで、特許全文を選択的に読むことができ、特許読解に要する時間を大幅に短縮することが期待できる。

2.3 特許情報の意味構造化

最近行われている、セマンティックコンピューティングと呼ばれる研究プロジェクトは、人とコンピュータが共有する意味に基づいてコンテンツを設計・運用しようというものである。設計の最初の段階から人間にも機械にも理解できる意味やボキャブラリーを使うので、ユーザーにとって意味のあるサービスを体系的に提供することができる。また逆に、ユーザーがシステムを理解・作成・改良することも可能になる。

そのような設計のためにオントロジを使う(図 III-7)。オントロジの主な構成要素は、概念階層と属性定義である。例えば市民の下位概念に非勤労者と勤労者がある、勤労者には教師、会社員などがあるというツリー構造が概念階層である。属性定義は、例えば非勤労者のインスタンスには扶養者という属性があって、その値は勤労者のインスタンスであるというものである。

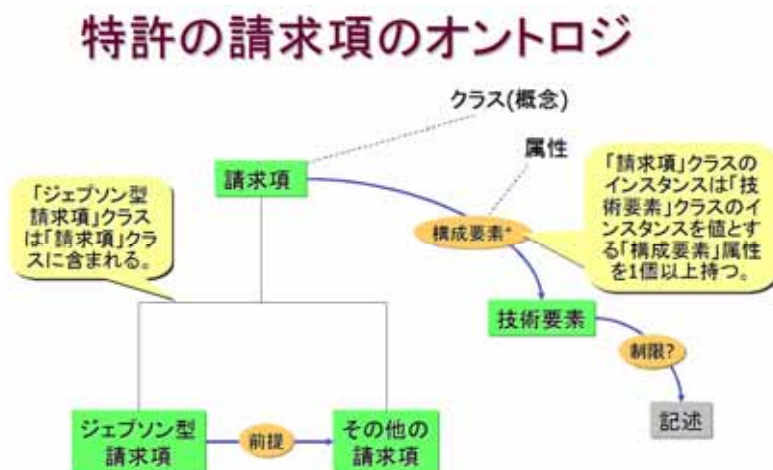
図 III-7 オントロジ



特許の情報もオントロジで定式化することができる(図 III-8)。請求項のサブクラスとしてジェプソン型とそうでないものがある。一般

に請求項は 1 個以上の構成要素から成立しているが、ジェブソン型の場合にはこの構成要素に加えて前提というものがある。前提というのは、ジェブソン型でない請求項の格好をしている。請求項をオントロジに従って構造化することにより、弁理士と発明者の間でのコミュニケーションが円滑になるとともに、要約や検索も容易になり、またフローチャートとしてビジュアルライズすること等もできる。翻訳の修正をする場合も、構造化されていれば、構造に基づいて翻訳対の部分同士の対応関係を簡単に把握できるので、10 倍以上の生産性の向上につながる。

図 III-8 特許の請求項のオントロジ



セマンティックオーサリングとは、オントロジに基づいてコンテンツの作成を支援する技術であり、コンテンツの作成段階から人手によって意味構造を明示しておくものである。既存のコンテンツに対して事後に構造化をするのは大変だが、コンテンツ作成段階から構造化しておけば、構造化の手間を考えると全体としてコンテンツの作成のコストが低減し、かつ、コンテンツの品質が向上する。

普通の文書の場合には、単文がノードで、接続詞に相当する意味的關係がリンクであるようなグラフにおいて、因果関係や逆接等の関係を人間が直接設定する。テキストを細切れにしておけば、現在の技術でもこれを解析してより細かい構造を自動的に高い精度で得ることが

でき、検索、要約、翻訳などが非常に高い精度で提供できる。

グラフ型のコンテンツの作成は以前から発想支援システムにおいて使われていた。グラフ型のコンテンツをまず作ってそれを見ながら文章を書くと良いコンテンツができる。良いというのは2つの意味があり、一つは見落としが少なく、つまり関連のある論点をより多く含むということであり、もう一つは考えが深まりやすい、つまり推論のチェーンが長いということである。しかし、従来の発想支援システムは意味的な関係が規格化・共有化されていなかったため、まとまった意味内容なり事柄を他人に伝えるための手段たり得ないという決定的な難点があった。

セマンティックオーサリングではまず意味的な関係のセットを規格化することを行っている。実際、ISOで国際標準化が進められており、2006年の8月ぐらいに素案ができる予定である。

特許の明細書は発明の詳細な説明と請求項の部分とから成っており、発明の詳細な説明は利用分野と課題と手段と実施例と効果と作用とから成っている。請求項は発明の詳細な説明のうち手段の部分に対応する。

以前、ジャストシステムと松下電器の間で係争になった松下電器の特許は、ある特定のアイコンをクリックした直後に別のアイコンをクリックすると後者のアイコンの機能が説明されるという、シンプルなものである。

これをオントロジに基づいて構造化してみると、手段の記述は明細書の中の様々なところで共有されていることがわかる。特に、請求項と発明の詳細な説明の手段のテキストとは微妙に言い回しが違うが同じことを言っているため、オーサリングの段階では共有できる。

請求項1および3には詳細な説明の手段の部分と全く同じ内容が共有されている。ほとんど手段で言っていることを繰り返しているにすぎないので、オーサリングの段階ではそのような繰り返しは不要であり、繰り返すかわりに共有すれば良い。共有しておけば修正等も共有される。手段のところを修正したのに請求項を修正し忘れるというこ

とはなくなるし、請求項の間での整合性の管理も簡単になる。

文書のいろいろな部分の間の対応関係を管理するのは非常に面倒な作業であるが、最初からコンテンツの部分を共有するというやり方でその対応関係を維持しておけば、作成作業はずっと楽になる。しかも、作成すべきボリュームも少なくなり、さらに構造化されているので、様々なアプリケーションに展開できる。

そのような構造化をせずに、既存のデータを自動解析して、その意味構造を使って検索している例を紹介する。「ロボットで住宅を作る」という検索質問文を入れて検索ボタンを押すと、約 10 万件の新聞記事のデータから様々な候補が出てくる。「作る」という用語の類語が並んでいるが、「建てる」「建設」「建築」などが上のほうに現われている。「住宅を作る」という意味的な関係をシステムが自動的に考慮してくれた結果である。それらで類語拡張して検索し直すと、正解が得られる。既存の文章を自動解析した精度の低い構造化であるにもかかわらず、その構造を手がかりとして類義語の選択を支援することにより、キーワードのみによる検索の場合と比べて、必要な時間は半分ぐらいになり、クリックの回数も 4 分の 1 ぐらいになるという予備実験の結果が得られている。

市販の文書作成ソフトには要約機能を持つものがあるが、あまり使われていない。精度が悪いからである。文章を初めから構造化しておけばはるかに安定して品質の高い的確な要約ができる。