

エンタープライズサーチ・エンジン QuickSolution® の開発

武 並 佳 則*・岸 田 正 博・田 辺 泰 夫

Development of Enterprise Search Engine “QuickSolution”—— by Yoshinori Takenami, Masahiro Kishida and Yasuo Tanabe —— As document digitization and information sharing increase in enterprises, the volume of information within a company’s possession grows steadily. As a result, there are increasing needs to effectively search across a company’s information assets. Enterprise search is a system for searching through all data owned by a company in a cross-sectoral manner. This paper describes the functions and features of the “QuickSolution” search engine developed as enterprise search software. First, five requirements (search result ranking, scalability for large-volume data, database search, access privilege management and customization) required for enterprise search are described, as well as the functions and the means of problem solution QuickSolution offers for each of these requirements. Next, more detailed descriptions are given on the two among the five requirements, which are the search algorithm related to search result ranking and the access privilege management function that is especially important for enterprise search.

1. 緒 言

文書の電子化による情報共有が急激に進み、蓄積されたデータ量は人が管理できる限界を超えている。これらの膨大な情報を効率よく検索したいというニーズは高まる一方である。GoogleやYahoo!のようなインターネット検索の飛躍がそれを象徴している。企業が所有する情報も日々増加しており、企業内の情報もGoogleやYahoo!のようにどこに何があるかを意識せず高速かつ精度よく検索したいというニーズが強くなっている。エンタープライズサーチとは、企業内の点在するデータを横断的に検索するためのシステムのことである。

本論文では企業向け検索ソフトとして開発したエンタープライズサーチ・エンジンQuickSolutionの機能・特長について説明する。まず、エンタープライズサーチに求められる5つの要件（検索結果ランキング、大容量データに対する拡張性、データベースの検索、アクセス権限管理、カスタマイズ）について説明し、各々の要件に対してQuickSolutionで提供している機能および課題解決の手段を述べる。次に、5つの要件の中で、検索結果ランキングに関わる検索アルゴリズム、および企業内検索では特に重要な要件であるアクセス権限管理機能について詳細を説明する。

2. エンタープライズサーチ・エンジンQuickSolution

2-1 エンタープライズサーチに求められる要件 企業内にエンタープライズサーチを導入するためにはインター

ネットの検索とは異なるいくつかの要件が存在する。本節では、その要件について説明し、各々の要件に対してQuickSolutionで提供している機能および課題解決の手段を述べる。なお、QuickSolutionはいくつかの製品群から構成されているが、本論文ではこれらのQuickSolutionシリーズをまとめてQuickSolutionと呼ぶことにする。

(1) 検索結果ランキング

インターネットの検索ではHTMLファイルが主な検索対象であり、ハイパーリンク情報が存在する。Googleでは、PageRankと呼ばれるハイパーリンク前提のランキング技術により高い検索精度を維持している。

一方、企業内の検索ではファイルサーバやデータベース・サーバ等、一般的にはハイパーリンクが存在しない。そのような状況で大容量のデータを効率的に検索するためには、ハイパーリンクに頼らないランキング技術が必要となる。QuickSolutionでは統計的な手法等の様々な技術により検索結果のランキングを行っている。QuickSolutionの検索アルゴリズムについては3章で詳細を説明する。

(2) 大容量データに対する拡張性

ファイルサーバ、Webサーバ、データベース・サーバ、各種グループウェア等の増大し続けるデータに対し拡張性を確保する必要がある。QuickSolutionでは1台のPCサーバで1テラバイト（TB）、1億件という大容量データを検索できる。また、1TBを超えるデータに対しては、複数台の検索サーバによる分散検索機能で検索を行うことができる。

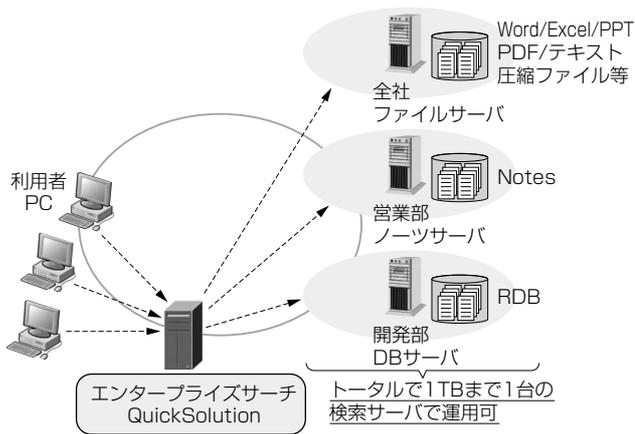


図1 エンタープライズサーチの概念図

(3) データベースの検索

基幹系をはじめとする業務システムの多くはリレーショナル・データベース（RDB）を利用して構築されており、そこに膨大な業務データが格納されている。エンタープライズサーチではこれらの業務データも検索対象になりうる。QuickSolutionでは、RDBはもちろんのこと、XMLデータベースも検索できる。データベースの検索では項目ごとの個別の検索に加え、項目をまたがる全文検索、特定項目を重視した検索等、充実した検索機能を提供している。

(4) アクセス権限管理

インターネット検索と違い、企業内検索ではアクセス権限の考慮が必須である。QuickSolutionでは充実したアクセス権限管理機能を提供している。具体的にはActive Directory連携、LDAP連携や、各種検索対象の権限継承機能等である。アクセス権限管理機能については、4章で詳細を説明する。

(5) カスタマイズ

企業内でエンタープライズサーチを活用するためには、基幹系システムとの連携においてセキュリティ要件に対応する等、様々な要件に対応するため、柔軟にカスタマイズできることが重要である。QuickSolution本体はJavaで実装しており、Java RMIのAPIを使用して柔軟なカスタマイズが可能である。また、REST形式のWeb APIも提供しており、様々なプログラミング言語からQuickSolutionを利用することができる。Web APIを使用することにより、WebサービスとしてQuickSolutionを利用すること、他のWebサービス、他システムとも容易に連携でき、利用実績も多い。

2-2 QuickSolutionの機能概要 QuickSolutionの概念図を図2に示す。

QuickSolutionは他のサーチ・エンジンと同様、対象データを読み込み、検索のためのデータ構造である独自のインデックスを生成する。このインデックスを使用することにより高速かつ高度な検索を行うことができる。

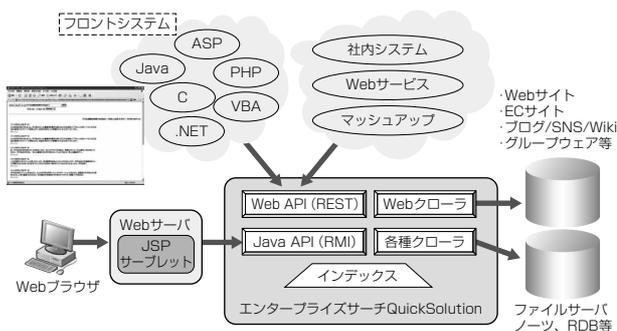
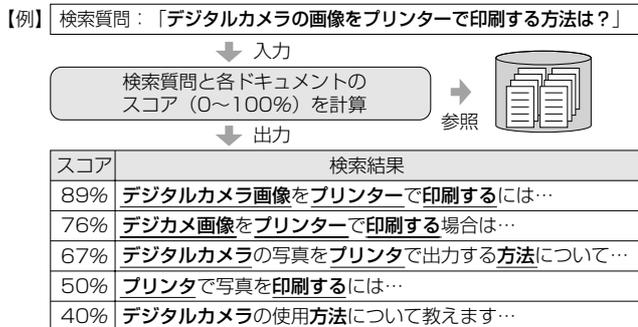


図2 QuickSolution概念図

QuickSolutionは以下の3種類の検索機能を提供している。

- ・全文検索機能
 - 類似検索（自然文によるあいまい検索）
 - キーワード検索（キーワードを指定した完全一致検索）
- ・属性検索機能（数値型、日付型等のフィールド検索）

上記の中で類似検索は特にユニークな機能で、図3の例のように自然文によるあいまい検索ができる。辞書登録等を行わなくても「デジタルカメラ」と「デジカメ」、「プリンター」と「プリンタ」等の表記の揺れを吸収して検索することができる。検索結果は表記揺れも考慮してスコア順にランキング表示される。



「デジタルカメラ」と「デジカメ」、「プリンター」と「プリンタ」等、表記の揺れを吸収。

- ▶ キーワードやフレーズでも類似検索が可能
- ▶ キーワード全文検索を圧倒する回答力を実現

図3 類似検索の例

図4に検索画面例を示す。検索結果では、検索質問に含まれるキーワードが最も集中して出現するサマリ部分が一覧表示される。これにより目的の文書が見つかったかどうかを一目で確認できる。右側のキャッシュ表示では、検索質問に含まれるキーワードのハイライトの詳細を確認できる。

QuickSolutionの特長は表1の通りである。



図4 検索画面例（スタンダード画面）

表1 QuickSolutionの特長

特長	内容
大容量・高速	<ul style="list-style-type: none"> 1台のPCサーバで1TB、1億件のデータを検索可能 1TB以上のデータは分散検索で対応可能 1000万件のデータを0.1秒で検索可能
辞書メンテナンス不要	<ul style="list-style-type: none"> 辞書が不要なN-gram方式がベースであるため、検索漏れがない 独自の統計処理でノイズを抑制
多言語対応	<ul style="list-style-type: none"> 検索エンジンに加え、画面も多言語対応（日本語、英語、中国語） 韓国語も検索可能
連続運用が容易	<ul style="list-style-type: none"> インデックスの差分更新で連続運用可能（再生成不要）で大容量データで非常に有利 インデックスサーバ機能を使用することにより、複数の検索サーバによるアクセス負荷分散、高可用性確保
充実した権限管理機能	<ul style="list-style-type: none"> Active Directory連携（シングルサインオン可能）、LDAP連携 ファイルサーバ、Net-It、ノーツ、Exchange、GlobalDocの権限継承可能 リアルタイム権限継承
多彩なデータ形式に対応	<ul style="list-style-type: none"> RDB/XMLに強い 動的ページ・認証ページ対応のWebクローラ機能を使用することにより、ブログ、Wiki、SNS等の動的生成コンテンツも検索可能
分析系機能の充実	<ul style="list-style-type: none"> キーワード抽出、関連語抽出、可視化、クラスタリング、シソーラス構築支援機能等の分析系機能の提供
カスタマイズが容易	<ul style="list-style-type: none"> 純国産 Javaで実装、プラットフォームを選ばない APIが充実 Java API (RMI)、Web API (REST)

■ 文書集合に対する検索質問の結果

	検索された文書	検索されなかった文書
適合文書	w	x
非適合文書	y	z

■ 再現率 (recall) = $w / (w+x)$

⇒ 検索結果にどれだけ漏れがないか

■ 適合率 (precision) = $w / (w+y)$

⇒ 検索結果にどれだけノイズがないか

図5 再現率と適合率

3. 検索結果のランキング

3-1 情報検索システムの検索精度

一般的に情報検索システムの検索精度は図5に示す再現率と適合率で評価する必要がある。再現率と適合率の総合的な評価尺度として、11点平均適合率、F尺度等が提案されているが、これらの評価尺度を適用するためには、①検索対象となる文書集合、②検索質問集合、③各検索質問に対して適合する文書集合の対応表の3つの要素を持つテスト・コレクションが前提となっている⁽¹⁾。

企業内検索では上記評価尺度を適用するのは通常は困難である。また、インターネット検索と違い、企業内検索では検索漏れが致命的になる場合が多い。そこで、QuickSolutionでは、

- ・ 検索漏れを防ぐ（再現率向上）
- ・ 検索結果上位のノイズを削減する（適合率向上）

という観点でアルゴリズムの改善および機能拡張に取り組んでいる。なお、ここでは検索対象のファイルサーバの各ファイル、データベースの各データ（レコード）等、検索結果一覧に表示される検索結果の単位をまとめて「文書」と呼ぶことにする。

3-2 検索アルゴリズム

検索条件を入力し、検索結果の1~2ページ目（上位20件程度）に探している文書がピンポイントで表示されることが望ましいが、ありふれたキーワードを1つだけ指定した場合や、同じような文書が多く存在する場合、目的の文書の上位表示が難しい場合がありうる。その場合に検索漏れを起こさず、他の検索条件を併用して効率的に絞り込めることが重要である。

(1) 基本アルゴリズム

図3で示したように、QuickSolutionでは全文検索（類似検索およびキーワード検索）の結果をスコア順に出力することができる。具体的には図6のようなアルゴリズムで検索を行っている⁽²⁾。

図6で出現頻度df（document frequency）は文書集合に

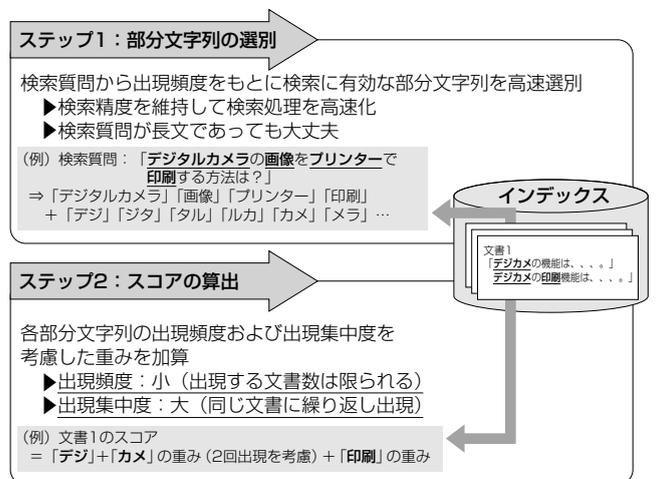


図6 類似検索アルゴリズム

において文字列（キーワード）を含む文書数である。また、出現集中度は文書集合において文字列を2回以上含む文書数df2を使ってdf2/dfで計算される統計量であり、ある文字列が文書に出現していることを条件として、その条件のもと出現回数が2回以上である確率の推定値となっている。

文書の主題を表すキーワードはある文書で一度使われると繰り返し使われるのが一般的である。出現集中度はこの性質に着目したもので、重要なキーワードの場合、出現集中度は大きな値をとる傾向にある。また、出現集中度は図7のようにキーワードの中では一定の値をとる性質があり、キーワードの識別・重み付けの決定に有用なことが分かっている(3)。そして、出現集中度を使うことはQuickSolutionの検索精度の良さに寄与している。

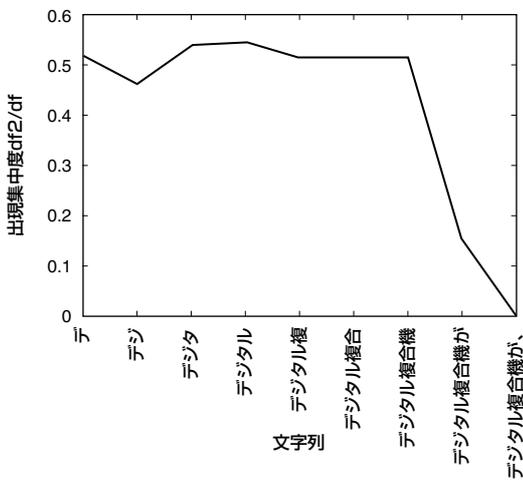


図7 出現集中度の性質

検索方式には大きく分けて、N-gram方式と形態素解析方式の2つの方式が存在する。N-gram方式は、システム辞書を使用せずに、検索対象のテキストを1~4文字ごとに区切りインデックスを生成する方式である。一方、形態素解析方式は、システム辞書を使用して、検索対象のテキストを単語ごとに区切りインデックスを生成する方式である。

QuickSolutionでは、ベースは辞書が不要なN-gram方式により検索漏れを防止している。すなわち、どのような分野のデータでも検索することができる。そして、統計処理に、出現頻度だけでなく独自の出現集中度を考慮したスコアを算出する。また、N-gramをベースに、形態素解析を併用した全文検索（ハイブリッド方式）も実装している。これにより漏れない全文検索を行うとともに形態素解析によるキーワードの有無をより重視した全文検索を行うことができる。

以上により、検索漏れを防ぐとともに（再現率向上）、検索結果上位のノイズを削減している（適合率向上）。

表2はN-gram方式、形態素解析方式、QuickSolutionのハイブリッド方式の比較である。ハイブリッド方式は、検索漏れがない、辞書のメンテナンス不要等のN-gramの長所をそのまま継承して、形態素解析の効果により検索結果上位のノイズを削減できる。

表2 検索方式比較

	N-gram方式	形態素解析方式	ハイブリッド方式
システム辞書	不要	必要	必要 (必須ではない)
インデックス形式	N-gram 転置ファイル :フルテキストインデックス	単語転置ファイル :単語インデックス	N-gram 転置ファイル +単語転置ファイル
長所	任意の文字列を検索可能 検索漏れがない	検索結果にノイズが入り難い	・任意の文字列を検索可能 ・検索漏れがない ・検索結果(上位)にノイズが入り難い
短所	検索結果にノイズが入り易い ・「京都」で検索して「東京駅」がヒット。 ・「スキー」で検索して「ウイスキー」がヒット	検索漏れが生じる →リカバーできず致命的 新語・造語が検索できない →辞書のメンテが必要 インデックス再生成が必要 ・型番の部分一致 ・「イントラ」で検索して「イントラネット」がヒットしない。	・インデックスサイズが大きい (従来比1.1倍程度) ・CPU、メモリの負荷は高い (インデックス更新時間は従来と同程度) ※検索漏れはN-gramベースでカバーできるため辞書のメンテおよびインデックスの再生成は必須ではない。

QuickSolutionでは検索質問が複数キーワードを含む場合にその網羅性を重視する検索、更に検索対象の各文書が長文の場合に、検索質問に含まれる複数キーワードの出現位置、出現集中度を考慮する位置考慮検索等も実装しており、検索精度を向上させている。

(2) 組合せ検索、多重ソート、アクセスランキング

上記アルゴリズムで検索を行った結果、ヒット件数が多い場合は絞り込みが必要となる。QuickSolutionでは検索漏れがないので、類似検索の結果を完全一致検索（ある単語を含む、含まない）や属性検索（日付等の範囲指定）との組合せ検索で効率的に絞り込むことができる（図8）。

検索結果から目的の文書に素早くたどりつくには、利用者が指定する並び順で検索結果の並び替えができることが望ましい。QuickSolutionではスコア順⇒日付順等の複数のソートキーを指定した並び替え機能（多重ソート）を提供している。また、一般に多くの利用者により参照される文書は重要な文書である可能性が高いと考えることができる。QuickSolutionでは検索結果において、各文書が利用者



図8 組合せ検索 (アドバンスド画面)

により参照された回数 (アクセス数) を記録して、アクセス数の多い順に検索結果を出力する、いわゆるアクセスランキング機能も提供している。

(3) 検索質問拡張

指定した検索質問に同義語、類義語等の関連語を付け加えて、新たな検索質問とすることを検索質問拡張という。検索質問拡張により、最初の検索質問に含まれていないキーワードを含む文書も検索することができ、一般的には再現率が向上する。QuickSolution では下記に示す検索質問拡張機能を提供している。

(a) キーワード抽出・関連語抽出による気付き支援

図9に示すように、検索結果の文書に含まれるキーワードを辞書を使用せずに統計的アルゴリズムで自動抽出することができる⁽³⁾。また、抽出したキーワードの中で、検索質問と同時に出現する確率の高いキーワードを関連語として抽出することもできる。これらのキーワード・関連語を検索条件に追加して絞り込むことができ、気付き支援機能として使用することができる。なお、本機能を用いて、絞り込み検索を行うことで適合率が向上する。



図9 キーワード抽出と関連語抽出

(b) 関連語辞書、シソーラス連携

QuickSolutionの類似検索では表記揺れを吸収して検索するが、「プリント」と「印刷」のような意味が同じで表記が異なる場合、単独では同一視して検索することはできない。この場合、関連語辞書に登録することにより、同一視して検索することができ、再現率が向上する。

また、企業名、自動車、医学用語等、13分野の完成された関連語辞書、すなわちシソーラスと連携する機能も提供している。

(c) シソーラス構築支援機能

本機能は、インデックス生成済みの検索対象に対し、統計処理を用いて自動的に関連語対 (シソーラス) を抽出する機能である。検索対象からキーワードを抽出し、キーワードの前後の文字列を調べて、同じように使用されているキーワード対を関連語対として抽出する⁽⁴⁾。表3に自動抽出した関連語対の例を示す。抽出した関連語対のリストを上記した関連語辞書の形式 (CSV形式) で保存し、検索時に使用することができる。

表3 自動抽出した関連語対 (シソーラス) の例

関連語1	関連語2
SCSI	USB
スクリーンセーバ	スクリーンセーバー
横書	縦書
ディスプレイ	モニタ
USB	スキャナ
文書	文章
スキャナ	デジカメ
スクリーンセーバー	壁紙
SCSI	スキャナ
アドバイス	質問
BMP	JPEG
オフライン	切断
キーボード	ケーブル
SCSIカード	SCSI接続
全角	半角
アップグレード	バージョンアップ
USB	ケーブル
EXE	exe
デジカメ	デジタルカメラ
テキストボックス	挿入

(4) 検索結果のカテゴリ分類、可視化クラスタリング

これまででは検索結果をリスト形式で表示するのが一般的である。しかし似たような文書が含まれる場合は、同じような文書がリスト表示され、探したい文書に簡単にたどりつくことができない。この問題を解決するため、最近では検

素結果を分析・分類して表示するいわゆるクラスタリング技術が実用化されつつある。自動分類してカテゴリごとに分かり易く表示することにより、利用者はピンポイントで目的の文書にたどり着くことができる。QuickSolutionでは下記に示す2種類の機能を提供している。

(a) カテゴリ分類

検索対象がカテゴリ情報を属性として保持している場合、QuickSolutionの属性検索フィールドとして設定することにより、検索結果を高速にカテゴリ分類し、カテゴリごとのヒット件数を集計する(図10)。利用者は検索結果一覧に探している文書が存在しない場合、カテゴリの分類結果を手がかりに検索結果を絞り込んでいくことができる。検索結果のヒット件数が数十万件の場合でも1秒以内に高速集計処理できるのが大きな特長となっている。

(b) 可視化クラスタリング

本機能では事前にカテゴリを設定する必要がなく、自動的に検索結果を分析・分類することができる(5)。図11では、各文書は球で表示され、文書に含まれるキーワードを調べて、共通するキーワードが多い文書を類似している文

書と判断して近くに配置する。更に、類似文書をトピックマップとしてまとめることができる。利用者はトピックマップのグループ名や各文書の関連性を手がかりにピンポイントで目的の文書にたどり着くことができる。

4. アクセス権限管理とセキュリティ対策

4-1 セキュリティ要件

エンタープライズサーチではセキュリティ要件として認証、認可、監査が要求される。QuickSolutionではこれらに対してユーザ認証機能、アクセス制御機能、検索ログ収集機能を提供している。

(1) ユーザ認証機能

エンタープライズサーチでは認証により許された利用者だけが検索システムを利用できる必要があり、QuickSolutionでは後述の4種類の認証方式に対応している。

(2) アクセス制御機能

QuickSolutionには利用者に許された検索対象や、文書だけを検索結果として表示できる機能があり、後述の2段階のアクセス制御機能を提供している。

(3) 検索ログ収集機能

QuickSolutionには、どの利用者が、どのような検索対象や文書を、どのようなキーワードにより、いつ検索したかのログを収集する機能がある。内部統制の観点からは、ログを収集しておくことも重要な要件であるが、QuickSolutionでは、この1週間でよく使われた検索キーワードの一覧等、検索ログを集計・レポートする機能も提供している。

4-2 認証方式

QuickSolutionでは、認証情報が格納される認証マスタとして何を利用するか等の認証方式として、下記の4種類を選択することができる。

(1) 独自認証方式

認証マスタをQuickSolution自身で管理する方式である。ユーザ、グループ情報を登録するためのツールを提供している。

(2) Active Directory (AD) 認証方式

ADを認証マスタとして利用する方式である。アクセス制御ではユーザ、グループはADに登録されているものをそのまま用いる。

(3) ADシングル・サイン・オン

AD認証方式の高度な認証方式として、クライアントPCのWindowsのログオンにより、QuickSolutionにも自動ログインするADシングル・サイン・オン機能を提供している。

ADシングル・サイン・オンの仕組みを図12に基づき次に説明する。①利用者はクライアントPCでADにログオンする。この時、Windows内部に認証情報が作成される。②ブラウザでQuickSolutionの検索画面を表示しようとする。③QuickSolutionは未認証であることを検知すると、ブラウザへ認証情報を要求する。④ブラウザは

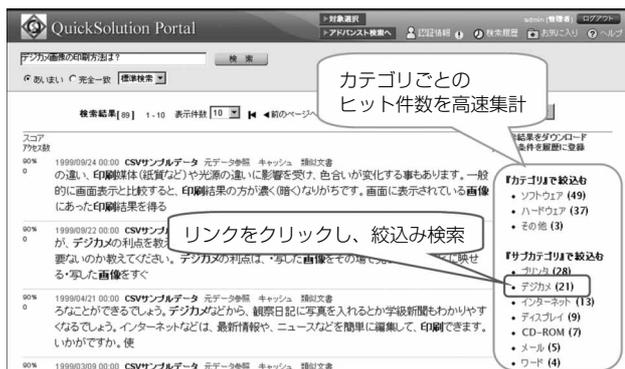


図10 カテゴリ分類

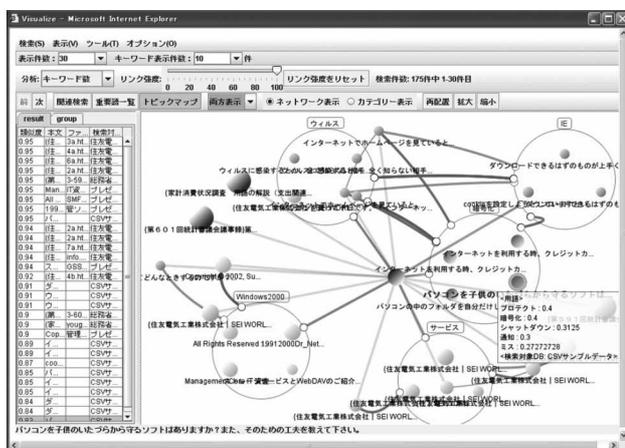


図11 可視化クラスタリング

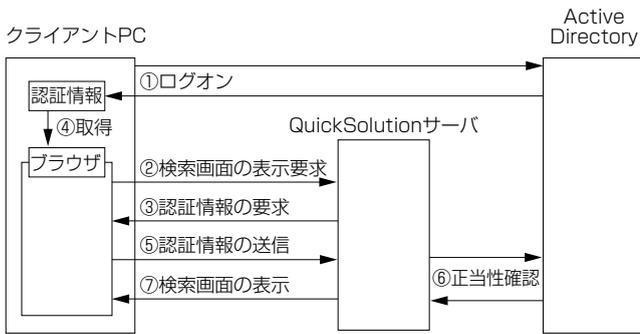


図12 ADシングル・サイン・オン

Windowsから認証情報を暗号化された状態で取得する。⑤ブラウザはこの認証情報をQuickSolutionへ送る。⑥QuickSolutionは受け取った認証情報をWindowsの機能を用いてADへ問い合わせその正当性を確認する。⑦正当性が確認されるとQuickSolutionは検索画面をクライアントPCに表示させる。

(4) LDAP認証方式

LDAP上のユーザエントリを認証マスタとして利用する方式である。アクセス制御で使うグループはユーザエントリ群を表すLDAPへの検索条件を使って定義することができる。

4-3 アクセス制御機能 QuickSolutionでは下記の2段階のアクセス制御機能を提供している。下記(1)、(2)はどちらか一方だけ使用することも、併用することもできる。

(1) 検索対象DBごとのアクセス制御

ファイルサーバごと、ノートDBごと等、ある程度まとまった検索対象DB単位でのアクセス制御を行う。図13に示すように各検索対象DBに対しユーザが所属するグループごとに、検索可/不可を設定することができる。

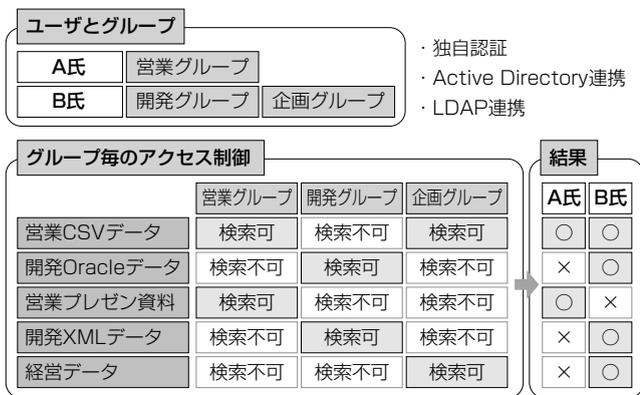


図13 検索対象DBごとのアクセス制御

(2) 文書単位の権限継承

検索対象DB内の文書単位に設定された閲覧権限を、そのまま継承して検索結果を表示することができる。例えば、ファイルサーバのフォルダ、ファイルに対し、ADのユーザまたはグループで権限が設定されている場合、検索結果一覧には自分の閲覧権限のあるファイルしか表示されない。

上述の仕組みを図14に基づき次に説明する。①利用者はクライアントPCのブラウザからQuickSolutionに検索を要求する。②QuickSolutionはインデックスから検索結果の候補を抽出する。③独自開発の権限チェックモジュールにより、結果候補から閲覧権限のあるファイルだけをフィルタリングにより得る。④QuickSolutionは1ページ分のファイルを検索結果としてブラウザに表示する。

権限継承はノーツ等のグループウェアに対しても同様に対応している。例えば、ノーツの場合、ノーツ文書だけでなく、ビューに設定されている権限も継承することができる。

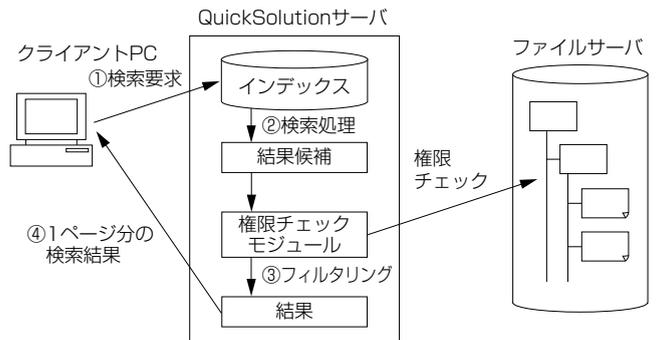


図14 権限継承の仕組み

4-4 リアルタイム権限継承 QuickSolutionでは急

な組織変更や閲覧権限の変更へ迅速に対応できるように、リアルタイムに権限のチェックを行っている。すなわち、ファイルサーバ側のADによる権限が変更されると、インデックスの再生成等を行わなくても、即座に検索結果に反映される。誤って権限のないデータ、ファイルが検索結果に表示されることはないため、情報漏洩も防ぐことができる。なお、検索実行時にリアルタイムに権限をチェックすると一般には検索速度が劣化するが、QuickSolutionでは権限チェックを検索結果の1ページ目に表示する数十件に絞る等の工夫により検索速度劣化を最小限に抑えている。

5. 結 言

エンタープライズサーチに求められる要件について説明し、各々の要件に対する QuickSolution の機能および特長を説明した。また、要件の中で、特に検索ランキング・アルゴリズムおよびアクセス権限管理機能について詳細を述べた。

今後も検索結果のクラスタリング機能の強化等、検索精度・速度向上に向け、検索アルゴリズムを強化する予定である。また、アクセス権限管理機能については階層的な管理者機能等の機能強化を行い、益々厳しくなるセキュリティ要件に対応できる製品に仕上げていく。

なお、QuickSolution は、IPA（現在の独立行政法人情報処理推進機構）次世代アプリケーション開発事業の中で、当社が豊橋技術科学大学と共同開発した類似検索技術を使用している。また、キーワード抽出およびシソーラス構築アルゴリズムは、IPA 未踏ソフトウェア創造事業の中で、豊橋技術科学大学梅村恭司教授が開発した技術を使用している。そして、可視化クラスタリング技術は、東京大学美馬秀樹准教授が開発した技術を使用している。紙面を借りて関係者に感謝申し上げます。

注 釈

・ PageRank :

Google 社が開発し、自社の検索エンジンに搭載している、Web ページの重要度の判定技術。「多くの良質なページからリンクされているページは、やはり良質なページである」という再帰的な関係をもとに、ページの重要度を計算している。

・ Active Directory :

マイクロソフト社によって開発されたディレクトリ・サービス・システムであり、Windows 2000 Server から導入されたコンポーネントである。

・ LDAP :

Lightweight Directory Access Protocol。ディレクトリ・サービスに接続するために使用されるプロトコルの一つ。

・ Java RMI :

Java Remote Method Invocation API。Java で書かれたプログラム間の ORB（オブジェクトリクエストブローカー）であり、RPC のオブジェクトに相当する機能を果たすための Java アプリケーションプログラミングインタフェース。

・ REST :

Representational State Transfer。元々は幅広い意味を持つが、ここでは Web API の一つで、URL パラメータに検索条件を与え、レスポンスに XML 形式のデータを受け取る HTTP リクエストを指す。

※本論文に記載のある会社名、製品名等は各社の登録商標もしくは商標です。

参 考 文 献

- (1) 徳永健伸、「情報検索と言語処理」、東京大学出版会（1999）
- (2) Eiko Yamamoto, Masahiro Kishida, Yoshinori Takenami, Yoshiyuki Takeda, and Kyoji Umemura, "Dynamic Programming Matching for Large Scale Information Retrieval", Proceedings of the sixth International Workshop on Information Retrieval with Asian Languages, July 2003, pp.100-108.
- (3) 武田善行、梅村恭司、「キーワード抽出を実現する文書頻度分析」、計量国語学第 23 巻第 2 号、pp.65-90（2001）
- (4) 當間雅、折原幸治、塩入寛之、梅村恭司、「関連語対のマイニングのための評価尺度」、言語処理学会第 13 回年次大会（2007）
- (5) Hideki Mima, Sophia Ananiadou, Katsumori Matsushima, Terminology-based Knowledge Mining for New Knowledge Discovery, ACM Transactions on Asian Language Information Processing, Volume 5, March 2006, pp.74-88.

執 筆 者

武並 佳則*：住友電工情報システム(株) ビジネスソリューション開発部
課長 部門スペシャリスト

岸田 正博：住友電工情報システム(株) ビジネスソリューション開発部

田辺 泰夫：住友電工情報システム(株) ビジネスソリューション開発部
チームマネージャー

*主執筆者