

「NAIST電子図書館学講座」
業務支援システムとそのデータベース
の構造

平成12年1月20日



株式会社リコー
ソリューション計画センター
平岡昭夫

hiraoka@src.ricoh.co.jp

内 容

1 .業務支援システムの概要

- システム設計上の留意点 (要求要件)
- 業務支援システムと電子図書館システムの連携

2 .全体システム構成

- ハードウェア構成、ソフトウェア構成
- G-BASE/FTS、アクセス権の設定

3 .データベース構成

- 目録DB、検索DB、ファイル構成

4 .利用者インターフェース

5 .全文検索システム

6 .LIMEDIO電子図書館オプション

システム設計上の留意点

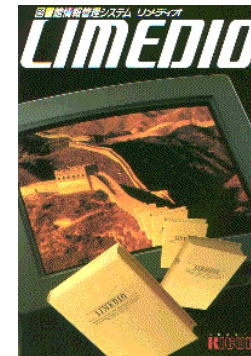
- 電子図書館はどこからでもいつでも誰にでも利用できること
- 画像などのマルチメディア情報にアクセスできること
- 一貫性のあるユーザーインターフェースを持つこと
 - (従来の資料を含む) 様々な資料を同じインターフェースから横断的に検索できる
- 従来の図書館サービスと電子図書館サービスが統合されていること
 - 従来の目録DBを電子図書館でも利用でき、OPACと電子図書館の検索が統合されていること
- 部分的な運用が可能であること
 - 書誌情報 → 目次情報 → ページイメージ → 全文テキスト
- 知的所有権を保護する機能を持つこと
- 規模の拡大やサービスの種類の拡大が容易であること

以上は統合化された電子図書館システムに対する基本的な要求要件

1 .業務支援システムの概要

業務支援システム

- リコー製「図書館業務パッケージ」を機能拡張して利用
 - 図書館情報管理システム LIMEDIO（リメディアオ）
- 目録管理システム、蔵書検索システムを中心に電子図書館システムとの連携機能を拡張
- セルフ貸出機能の追加、ビデオ・CD-ROM閲覧機能へのリンク設定などをカスタマイズ



図書館情報管理システムLIMEDIO

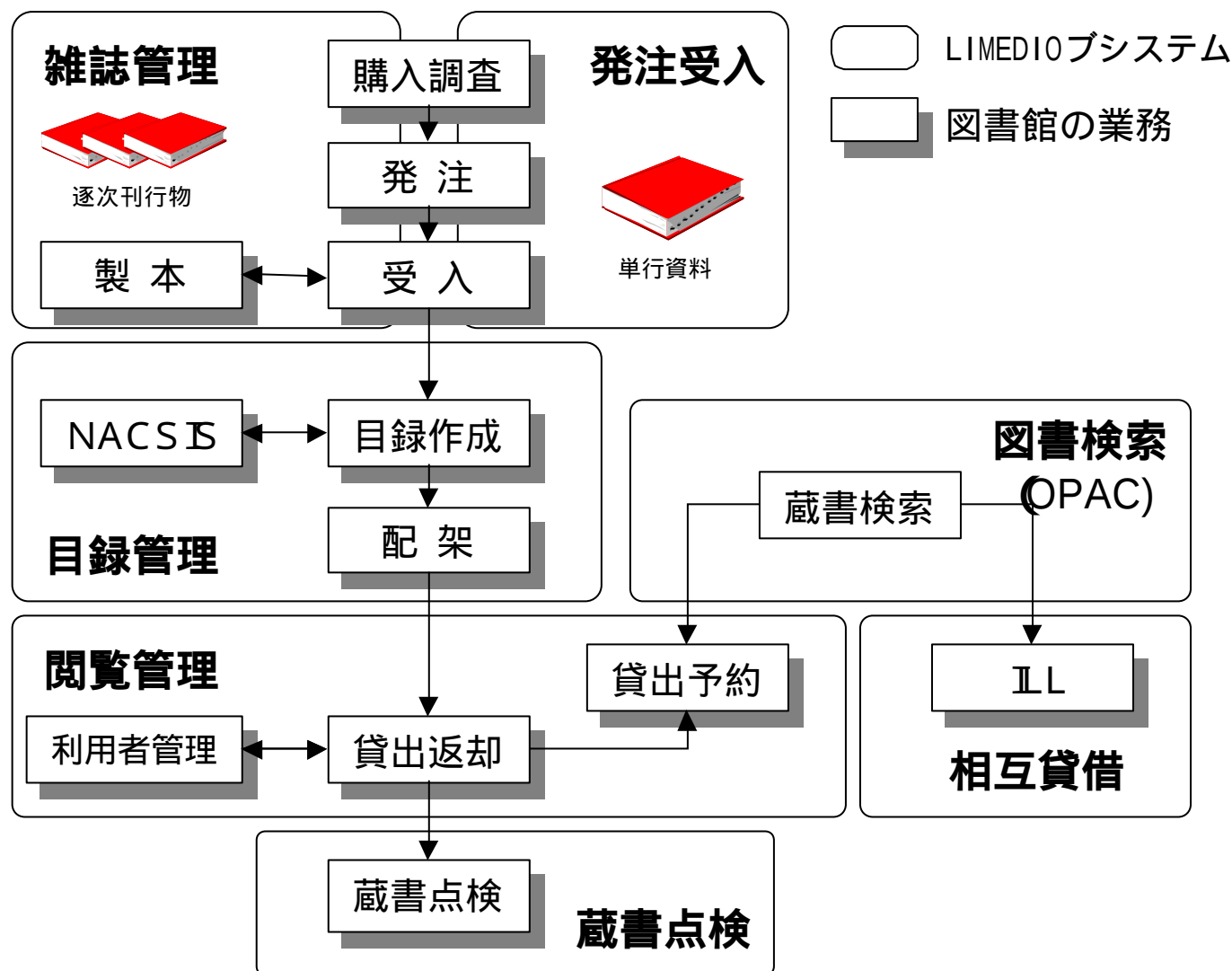
- 大学図書館向け標準ソフトウェア・パッケージ
 - 豊富な設定機能やオプションでほとんどの図書館の業務に適応可能
(導入実績 約 90校)
- UNIX サーバー、PC クライアントのオープンシステム
 - 大規模図書館 (分館対応) から小規模な資料室まで柔軟に対応可能
- 学術情報センター NACSIS-CAT/ILL にフル対応
 - 新 CAT プロトコルにも対応
- 先進技術を用いレベルの高い利用者サービス
 - 高速全文検索による「フリーキーワード検索」
 - Web 技術を駆使した WWW 版 OPAC
 - 貸出予約、購入依頼、複写依頼、予算照会 ... etc
- 特定メーカーにとらわれない柔軟なハードウェア構成

図書館情報管理システムLIMEDIO

基本システム (目録管理)と6つのサブシステムから構成される

基本システム	図書 雑誌目録管理、所在管理、配架処理、除籍処理 冊子目録、蔵書統計、雑誌統計、NACSIS-CAT連動
図書検索サブシステム	図書 雑誌目録検索、利用者予約、本人貸出照会 館内OPAC、学内LAN・インターネット対応
閲覧管理サブシステム	貸出返却、予約管理、延滞処理、督促処理、 罰則処理、利用者管理、利用統計、オフライン貸出
発注受入サブシステム	発注受入、登録処理、予算照会、 特別購入対応、各種帳票 (伝票)作成
雑誌管理サブシステム	雑誌(逐次刊行物)管理、購読調査、契約管理、 巻号受入、製本処理、一括受入、学総目データ作成
蔵書点検サブシステム	蔵書点検 (ハンディターミナル使用)
相互貸借サブシステム (ILL)	NACSIS-ILL 利用、ローカル処理 (受付、依頼、精算) 各種帳票作成、非NACSIS依頼対応、統計処理

図書館業務とLIMEDIOサブシステムの対応



基本システム : コード表管理、システム運用ツール等

電子図書館システムと業務支援システムとの連携

■ 入力業務

業務支援システムへの資料受入、目録登録（業）

資料の目次情報の作成、電子化一次情報の作成（電）

→ と の業務が並行して非同期に行える

■ 定期更新

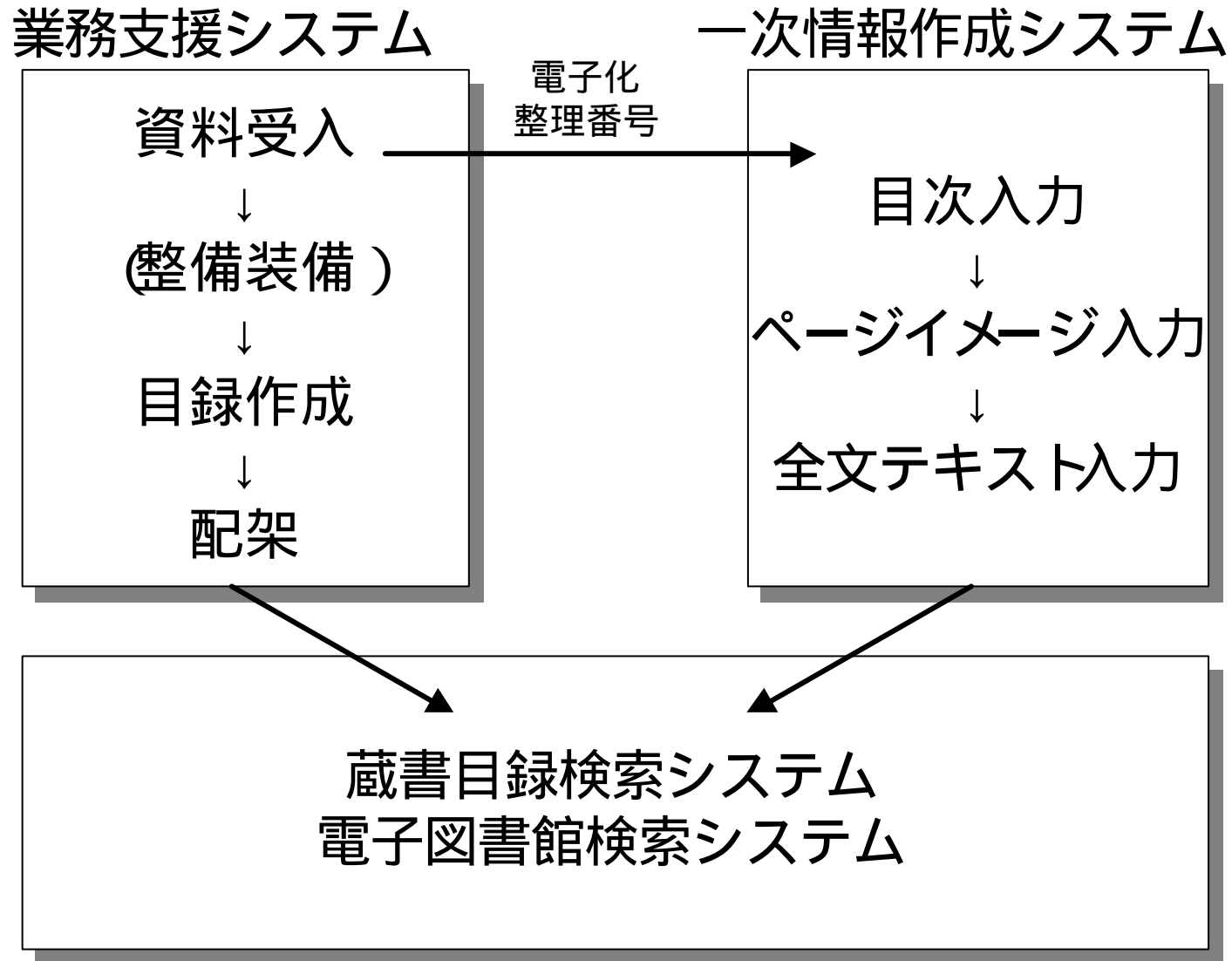
→ 業務用DBと検索用DBを分け、検索用DBへの定期更新を
入力業務が行われない深夜に行う

→ 図書館の開館時間に拘束されな検索閲覧サービス提供

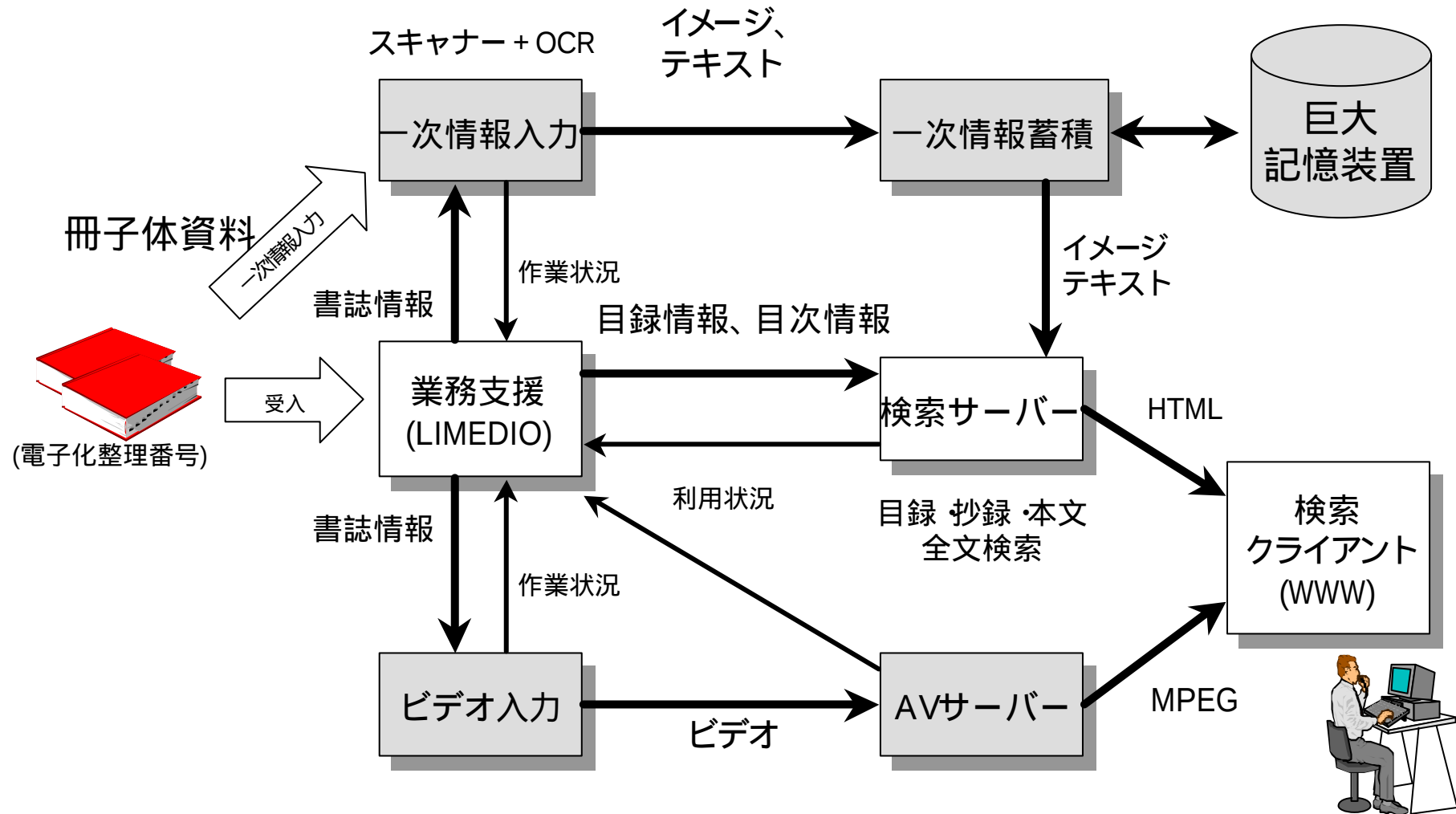
■ 検索閲覧

→ 資料に設定されているアクセス権に基づき、利用者の資料を
制限することで、著作権を保護する

資料電子化業務の流れ



電子図書館システム全体の流れ

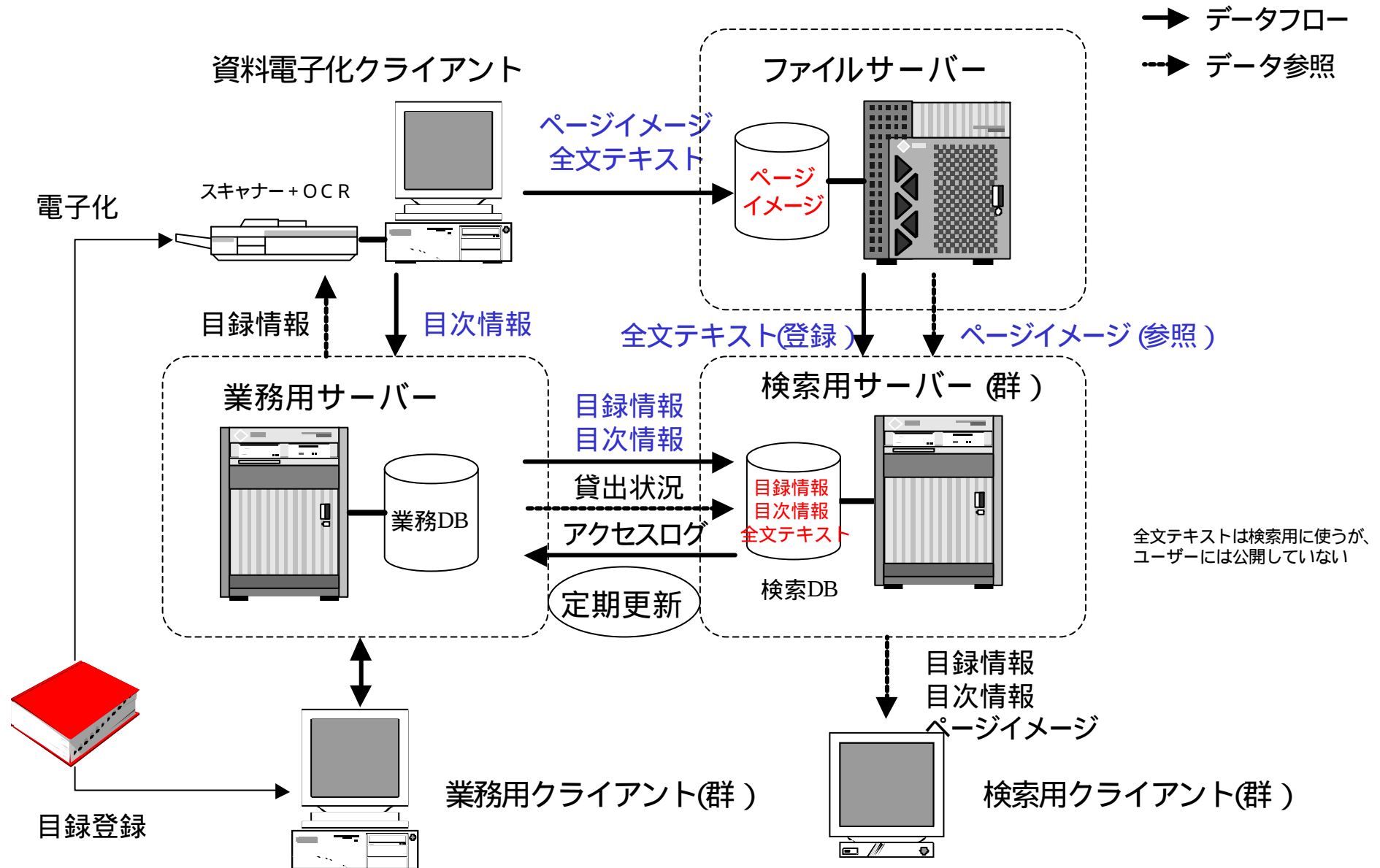


2.全体システム構成

全体システム構成の概要

- 業務サーバー、検索サーバー、ファイルサーバー
の3種類のサーバー
- 検索サーバーは5台の計算機による「協調分散システム」
- システムは「データ層」「ロジック層」「表示層」からなる3階層アーキテクチャー
- サーバーのOSはUNIX (日本語 Solaris7)、業務クライアントのOSは WindowsNT 4.0
- GUIは「X ウィンドウシステム (X11R5)」もしくは「Microsoft Windows」
→ CATPクライアントは Windows のみ
- DBMSは「拡張リレーショナルDBMS G-BASE」

システム構成 (ハードウェア)



分散アーキテクチャ概念図 (検索サーバー)

【検索サーバーの選択】

マスターサーバー (Dispatcher) は
各検索サーバーの負荷状況および
セッション数を管理している

クライアント (ブラウザ) からの接続
要求はマスターサーバーに一度集め
られる。

マスターサーバーは負荷状況に
より最適な検索サーバーを選択し、
スレーブサーバー (Dispatcher) にセッ
ション要求を出す。

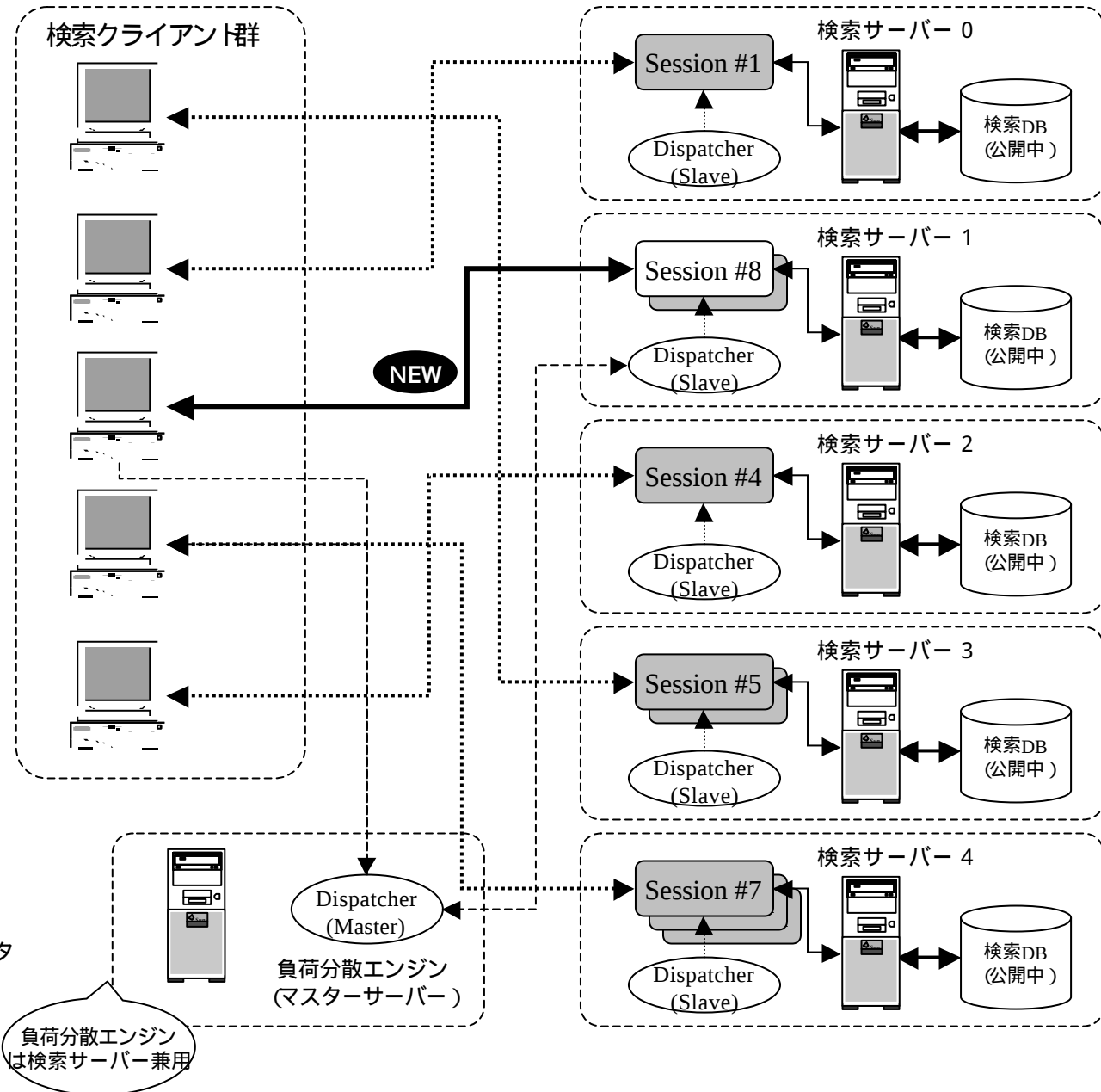
クライアントは Masterサーバーから
帰ってきたホスト+ ポート番号に
対してセッションをはり、以後の操作
を行う。

一度設定されたセッションの情報は、
ブラウザから開放要求が来るか、
タイムアウトするまで、それぞれの
検索サーバー上で保持される。

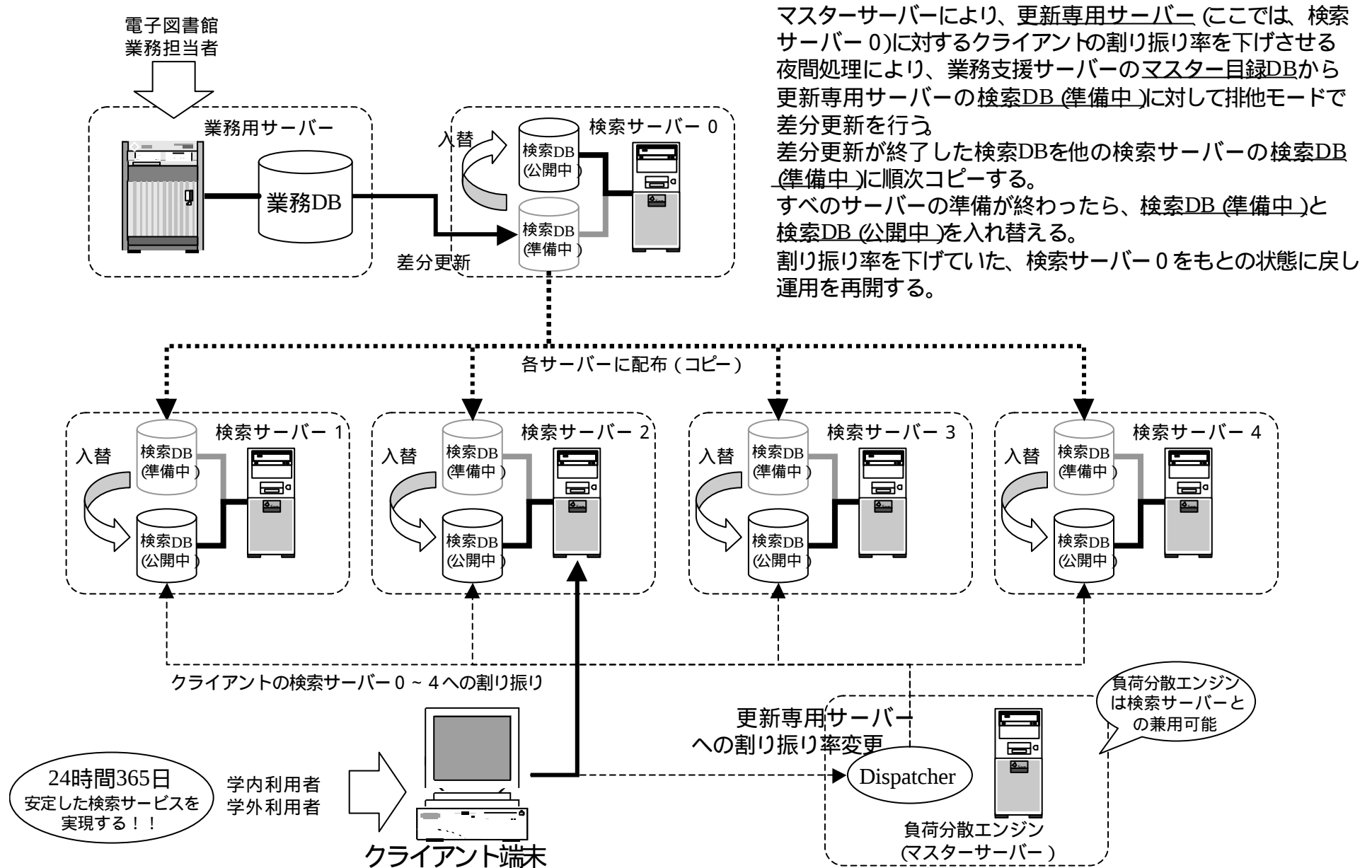
スレーブサーバーはセッションが開放
された情報を マスターサーバーに
通知する。

$$\begin{aligned} \text{負荷} &= W_{\text{load}} * \text{ロードアベレージ} \\ &+ W_{\text{nsess}} * (\text{セッション数/スケール単位}) \\ &- W_{\text{round}} * \text{選ばれなかった回数} \end{aligned}$$

W_{load} , W_{nsess} , W_{round} は加重のパラメータ

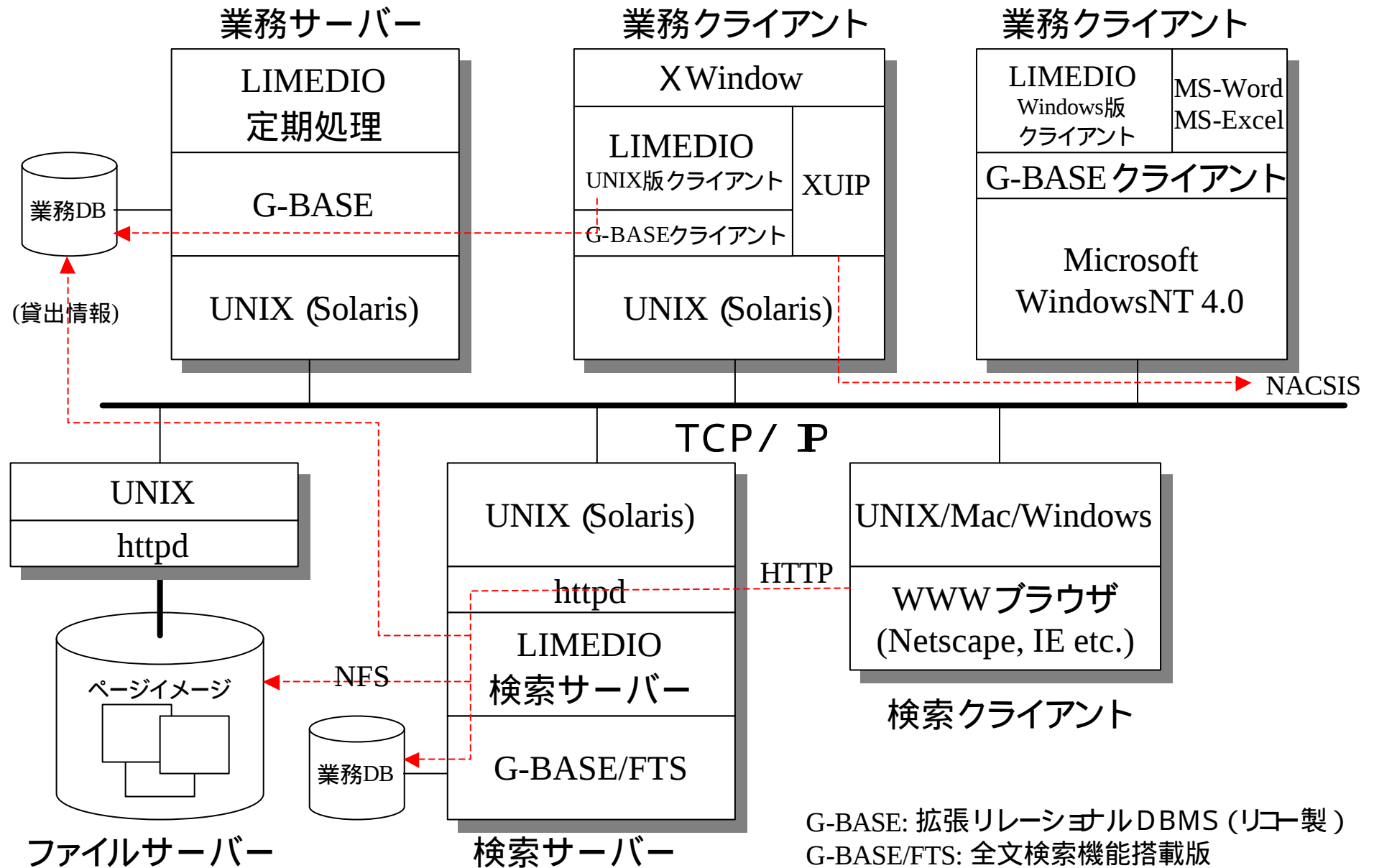


分散アーキテクチャ概念図 (定期更新)

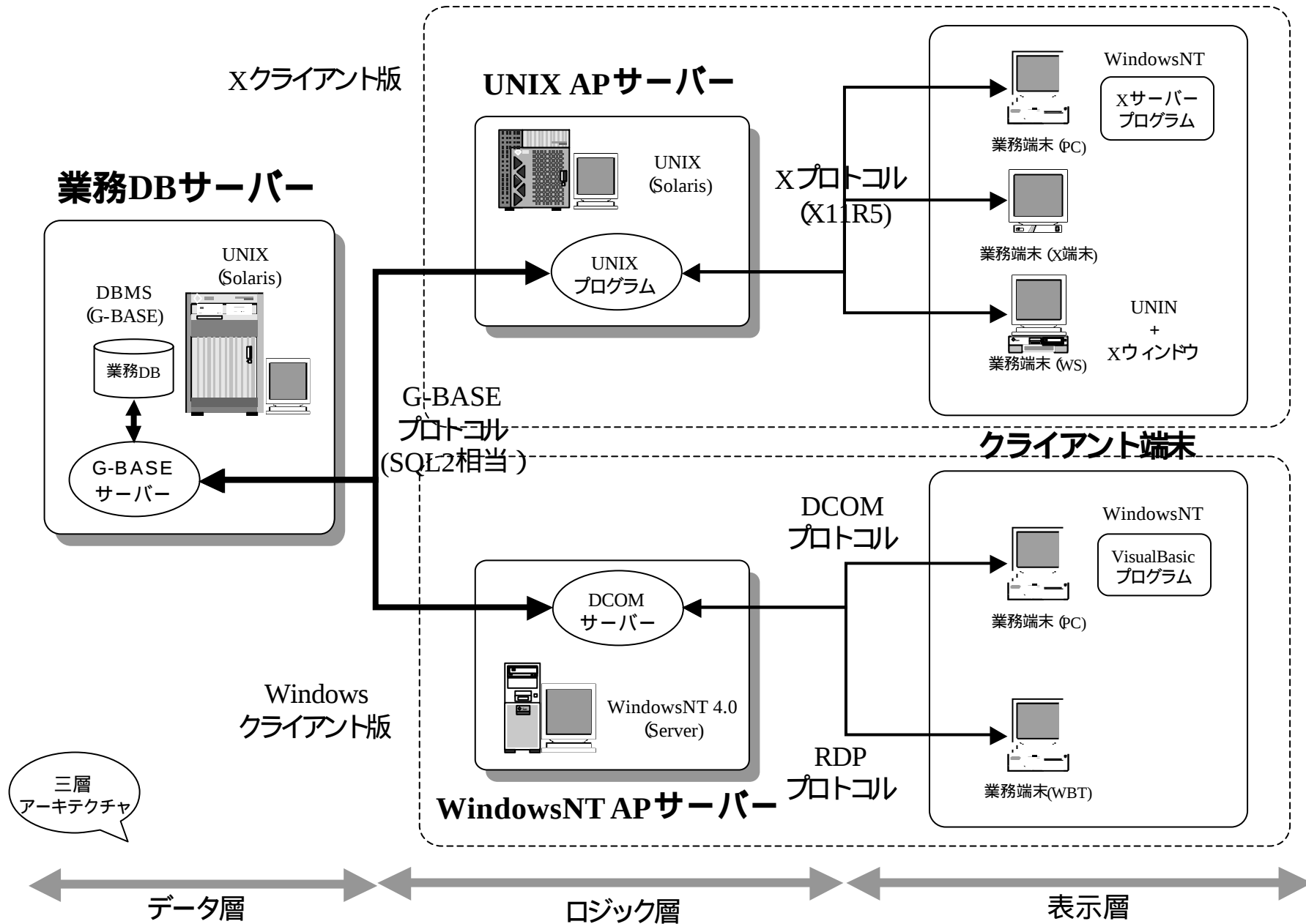


マスターサーバーにより、更新専用サーバー (ここでは、検索サーバー 0) に対するクライアントの割り振り率を下げさせる夜間処理により、業務支援サーバーのマスター目録DBから更新専用サーバーの検索DB (準備中) に対して排他モードで差分更新を行う。
差分更新が終了した検索DBを他の検索サーバーの検索DB (準備中) に順次コピーする。
すべてのサーバーの準備が終わったら、検索DB (準備中) と検索DB (公開中) を入れ替える。
割り振り率を下げていた、検索サーバー 0 をもとの状態に戻し運用を再開する。

システム構成 (ソフトウェア)



LIMEDIOのシステムアーキテクチャ (動作環境)



拡張リレーショナルDBMS G-BASE V3.5

- ERモデルを直接表現できるリンク機能
- 可変長データ 日本語処理に優れたRDBMS
- シグネチャーインデックスのより文字列の部分一致検索が高速
- FTS (Full Text Search) サーバー機能の搭載により全文検索機能 性能が大幅向上
- SQLインターフェース、ODBCインターフェース
- トランザクション処理、排他制御、アクセス権制御、リカバリー
- クライアントサーバー・アーキテクチャー
 - サーバー : UNIX、クライアント : UNIX, WindowsNT
- LMEDDに対応するため大幅チューニング

G-BASEは
リコーの独自開発
商品です



著作権の保護 (アクセス権の設定)

- ユーザーを「学外」「学内」「公開端末」の3種にわけてそれぞれアクセス権を設定
- 一般の電子化資料に対し以下の3種類アクセス権を設定可能
 - 自由にアクセス可能(F)
 - 図書館の業務支援システムの登録利用者のみアクセス可能(A)
 - 学内のネームサーバへ登録された利用者のみアクセス可能(H)
 - アクセス不可(X)
- 特定の電子化資料に対して個別のアクセス権を設定可能
 - 【例】「日経○○○」の雑誌だけは、一般資料と異なるアクセス権にしたい
- 学内外の判別は IP アドレス (サブネットマスク) を利用、公開端末はあらかじめホスト名を指定

例 1) 一般資料については、公開端末では自由に閲覧可能、学内からは図書館利用者であることを確認して可能、学外はアクセス不可 (XAF)

例 2) 特定の資料について、学内はフリー、登録ユーザーのみはユーザー ID とパスワードで本人を確認の上閲覧可能とする(HFF)

3. データベース構成

データベース構成の概要

■ 業務用マスターDB (群)と利用者用検索DBに分離

- 同じ目録DBでも「業務用」と「検索用」のではDBは構造が異なる
- 業務用マスターDBは、貸出返却発注・受入、ILL関係等の情報も管理

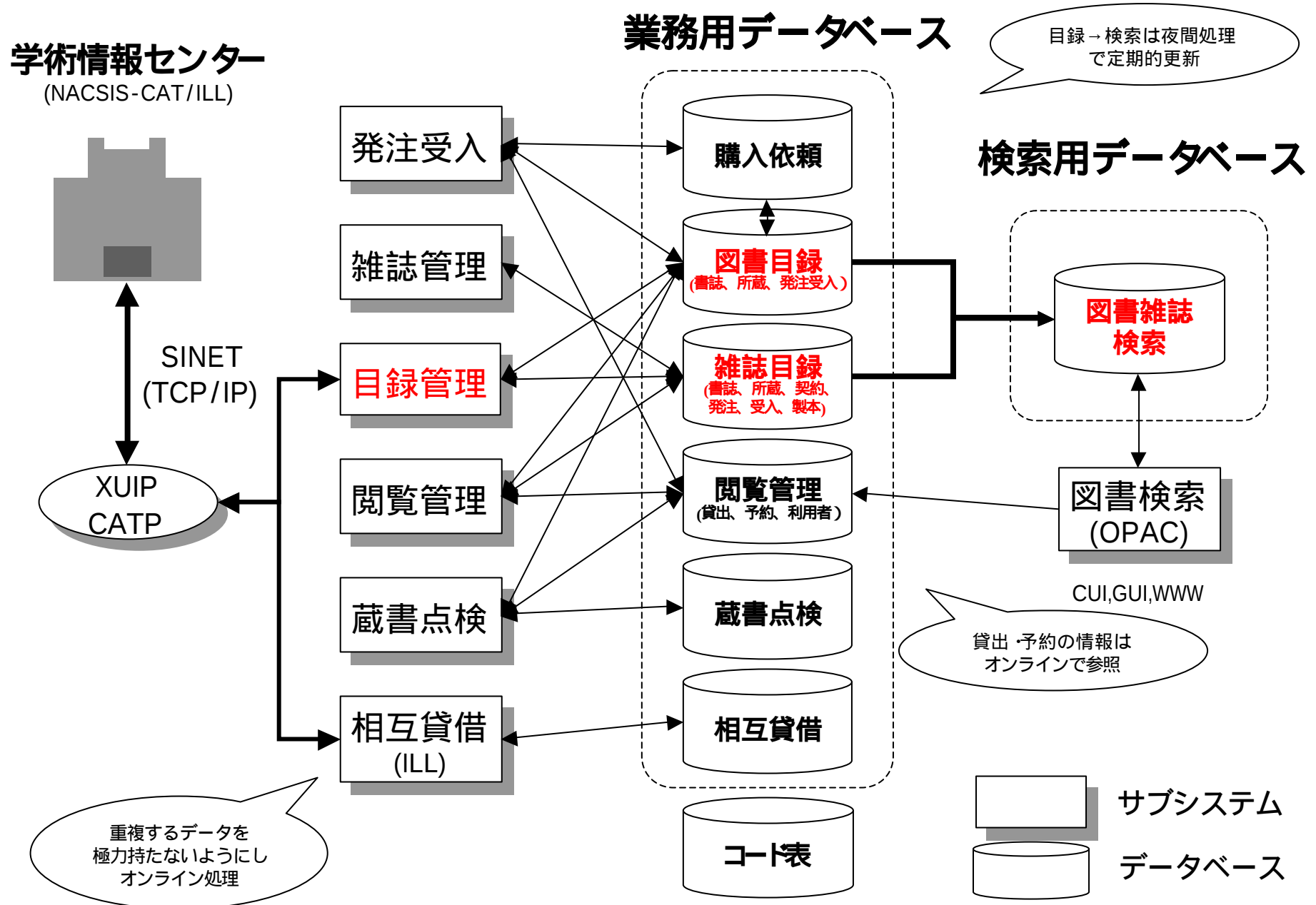
■ 目録DB (書誌・所蔵)は図書・雑誌とも3階層

- 図書：NC書誌、VOL書誌、所蔵
- 雑誌：雑誌書誌、雑誌巻号、コピー (受入・製本)
- 一次情報を管理するためには巻号レベルの書誌が必要

■ 電子図書館システムの書誌情報も目録DBで管理

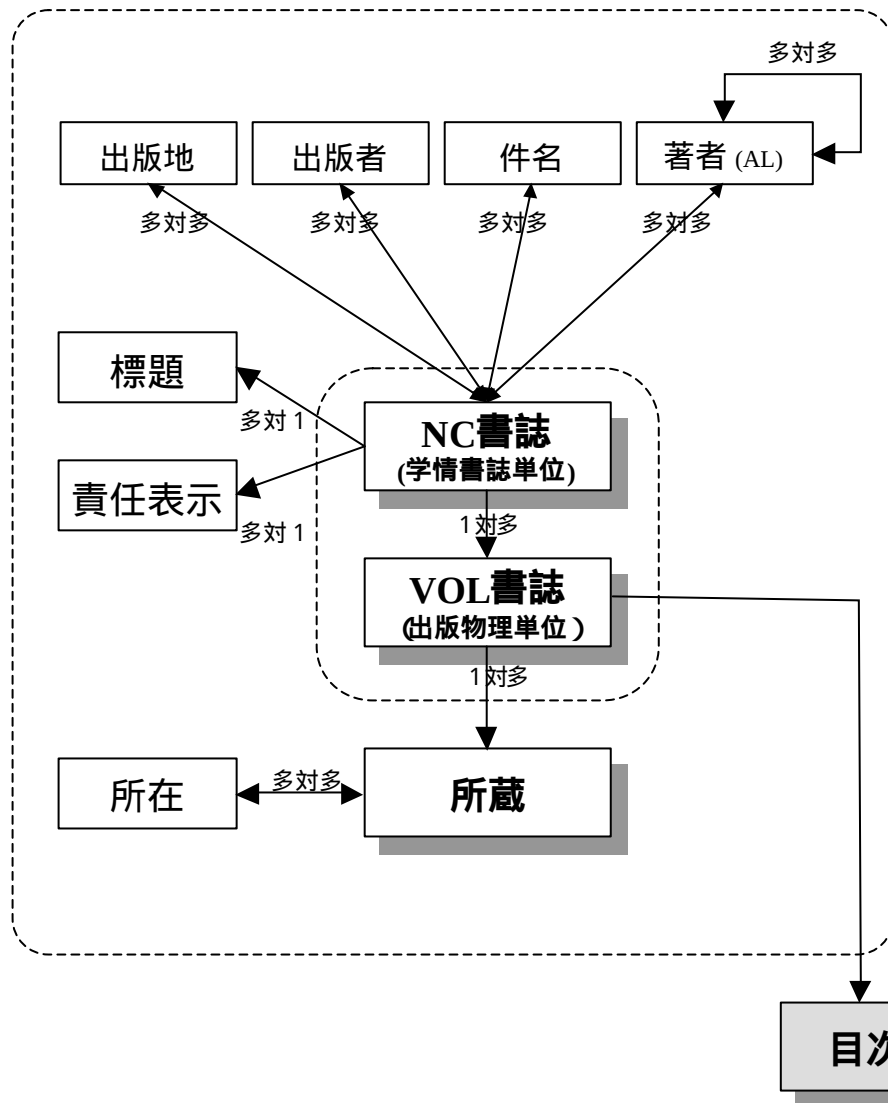
- (逆に)目録DBに登録されていない一次資料は管理できない
- 「電子化された資料 (情報)」と「冊子体で管理される資料」の両方を統一的に管理できる (検索できる)
- 「目次情報」「ページイメージ」「全文テキスト」など電子化のレベルが異なる資料の混在も許される
- 書誌・所蔵、目次情報は業務DB・検索DBの両方で管理。全文テキストは検索DBのみに格納
- ページイメージはファイルサーバーでOSが管理するファイルとして保存

業務支援システムのデータベース構造



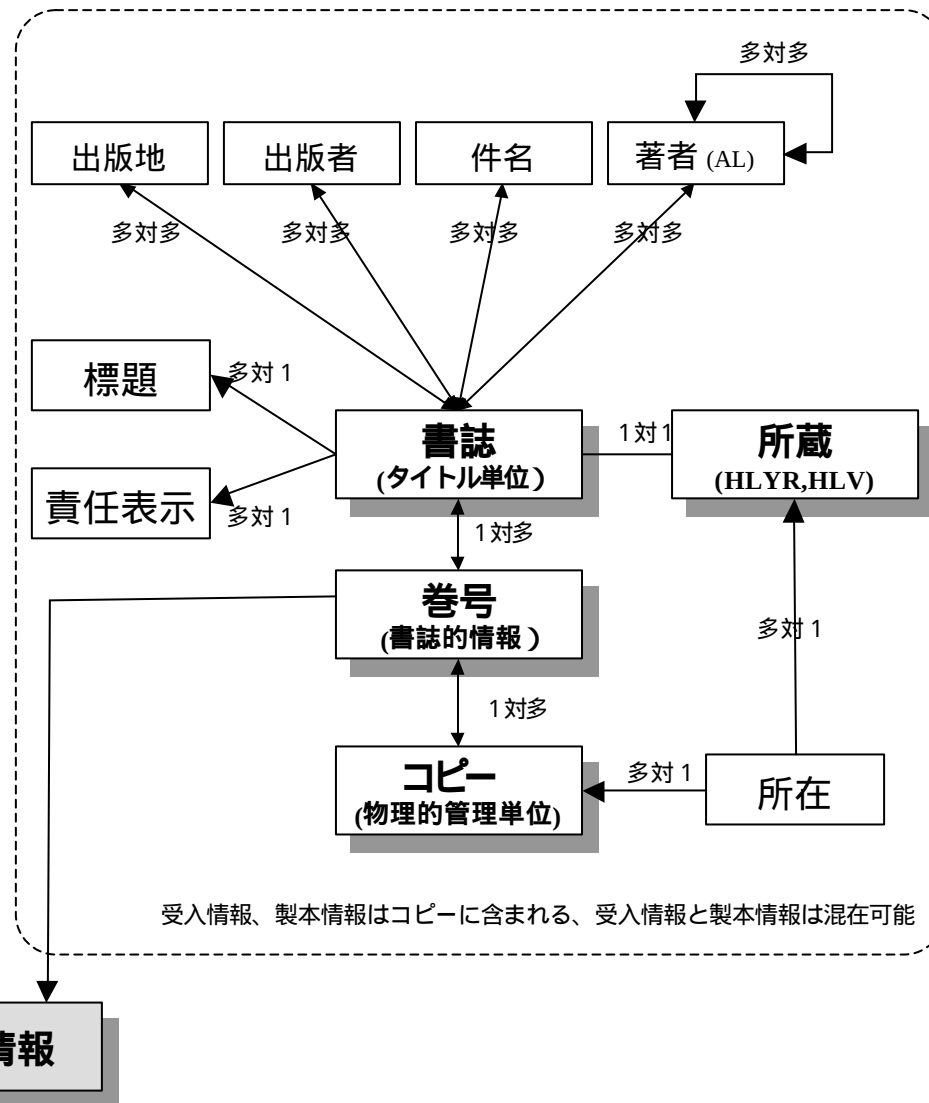
図書書誌データベース

NC書誌、VOL書誌をまとめて「分割書誌」とすることも可能（3つの構造を同一DBに併存可能）。

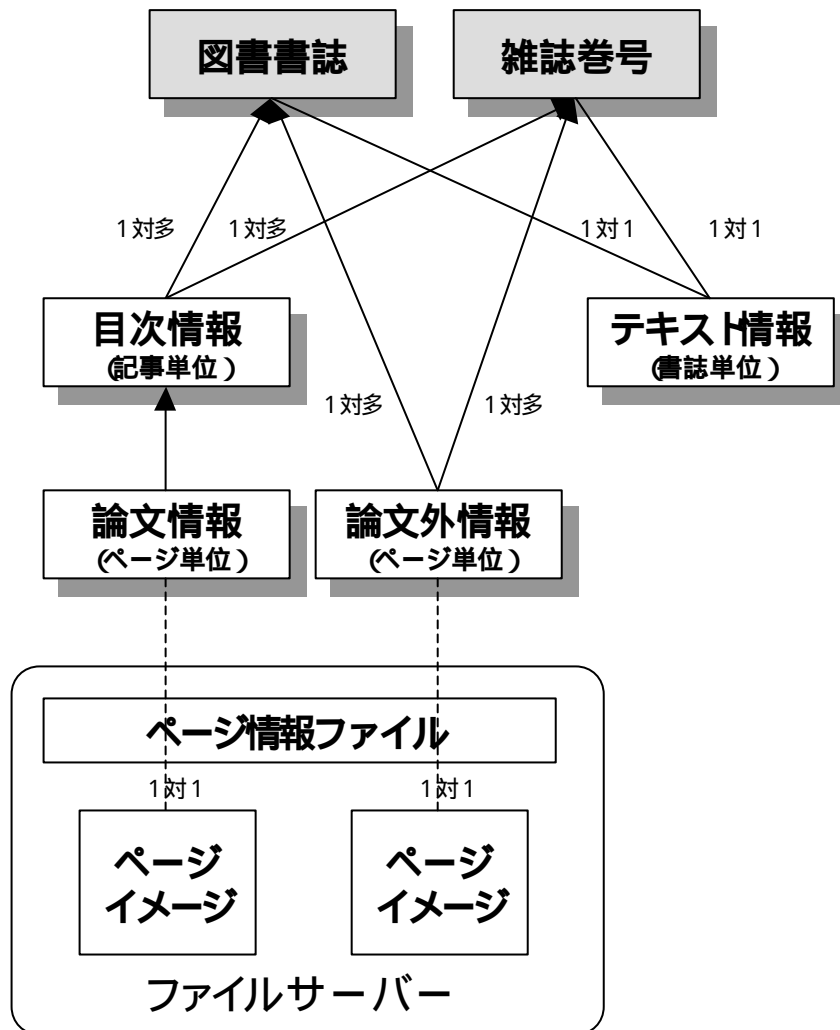


雑誌書誌データベース

「巻号」は巻号書誌情報、物理単位の情報管理は「コピー」
製本はコピーの一形態と考える



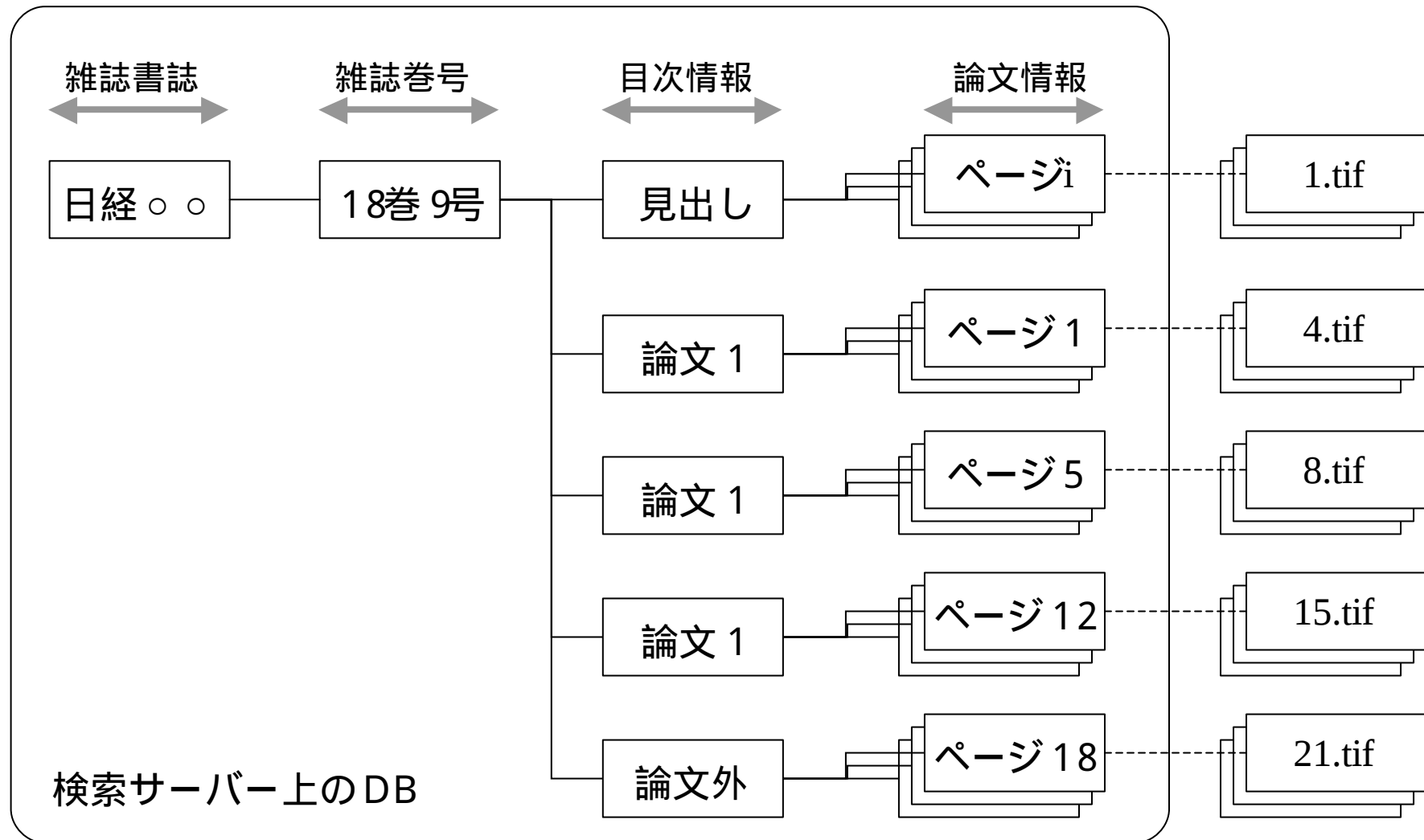
目次情報・全文テキストの管理



- 「目次情報」は記事単位に1レコード作成される
 - 標題、著者名、ページ範囲、キーワード、アブストラクト、メモ等を記録
 - 記事の内容を「論文(p)」と「見出し(h)」の区別が出来る
 - 一次情報の有無、アクセス権などを持つ
- 「論文情報」は論文や見出しをページ単位のレコードで管理
 - ページごとに全文テキストを保持
- 記事の形態を持っていない場合には「論文外情報」でページ単位のレコードのみを作る
- ページの概念がない場合、テキスト情報が直接書誌情報にリンクされる
- 格納されているテキストには全文検索用の索引ファイルが作成される
- 書誌ID (もしくは、巻号ID)とページ範囲の情報から該当するページイメージを得る。
- 「論文情報」「論文外情報」「ページ情報」は検索サーバーのみに存在する

目次情報・全文テキストの管理

【例】雑誌 1冊をすべて電子化した際の例



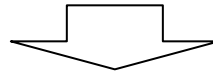
利用者検索の対象指定

利用者が検索対象をある程度しぼれるように検索対象を指定可能にした

書誌事項に現れる場合

アブストラクト及び目次情報に現れる場合

全文中に現れる場合



- 図書書誌 雑誌書誌 雑誌巻号を対象に検索
- に加え、目次情報 (記事のタイトル・著者名・抄録) を対象に検索
- に加え、論文情報 (ページ単位の全文テキスト) を対象に検索

ページイメージファイルの管理

- (原則として) 1ページに対して1イメージファイルを作成・管理する
- イメージファイルは UNIX 上のファイルとして保管される
- 図書書誌 雑誌巻号単位にディレクトリを作成
 - 図書: 'B' + 書誌Id (数字列)
 - 雑誌: 'M' + 巻号Id (数字列)
- 各図書書誌 雑誌巻号ディレクトリの下に ...
 - page: 表示用イメージファイル格納用サブディレクトリ
 - mini: サムネイル (縮小) イメージファイル格納用サブディレクトリ
 - orgn: オリジナルイメージファイル格納用サブディレクトリ
 - page.info: ページ情報ファイル
- サブディレクトリにはページ連番を名前とするイメージファイルを格納する
(表紙、目次、索引、裏表紙等も含めてページ連番を振る)
- ページ連番と実際の資料のページの対応を「ページ情報ファイル」に記述する

【例】図書書誌Idが 123456 の 3 ページ目の
ページイメージ情報 (表示はGIF形式、
元画像はTIFF形式の場合)

/opt/lime/pub/var/dlam の下に

```

/B123456 └─ image └─ page ─ 3.gif
                  └─ mini ─ 3.gif
                  └─ orgn ─ 3.tif
                  └─ page.info
  
```

4 .利用者インターフェース

利用者インターフェースの特長

- 標準検索インターフェースとしてWWWブラウザ採用
 - どこからでも誰でもアクセスできる
 - 各種マルチメディアをアクセスできる
 - 負荷分散を図ることができる
- 電子図書館システムと従来型蔵書検索システムを統合的に利用できる
 - 書誌情報の参照と一次資料の閲覧がシームレスに相互連携
- 高速で柔軟なフリーキーワード検索を対象資料を意識することなく利用できる
 - 提供対象は、図書書誌、雑誌書誌、巻号書誌、目次（論文）、アブストラクト、本文テキストで、横断検索も可能
- 「貸出予約」「購入依頼」「複写依頼」「状況参照」等をオンラインで行える

画面- 1) 検索条件の入力

図書資料検索

簡易検索

UMED3D Library Materials Search - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入りに登録(I) ツール(T) ヘルプ(H)

[操作方法ヘルプ / トップページへ](#)

検索

検索 検索条件クリア [検索キーごとの入力欄を表示しない](#)

- 検索対象: ☒ 図書 ☒ 雑誌名 ☒ 雑誌 ☒ 論文 ☒ 学内論文 ☒ ビデオ ☒ CD-ROM
- キーワード検索範囲:
- キーワード:
- タイトル:
- 編者名:
- 件名:
- 出版社:
- 出版年: から
- ISBN:
- ISSN:
- 新著: 日以内

検索 検索条件クリア [検索キーごとの入力欄を表示しない](#)

履歴表示

[操作方法ヘルプ / トップページへ](#)

Copyright (c) 1994, 1997, 1998, 1999 Ricoh Company, Ltd.
All rights reserved.

UMED3D Library Materials Search - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入りに登録(I) ツール(T) ヘルプ(H)

[操作方法ヘルプ / トップページへ](#)

検索

検索 検索条件クリア [検索キーごとの入力欄を表示する](#)

- 検索対象: ☒ 図書 ☒ 雑誌名 ☒ 雑誌 ☒ 論文 ☒ 学内論文 ☒ ビデオ ☒ CD-ROM
- キーワード検索範囲:
- 検索条件:

検索 検索条件クリア [検索キーごとの入力欄を表示する](#)

履歴表示

[操作方法ヘルプ / トップページへ](#)

Copyright (c) 1994, 1997, 1998, 1999 Ricoh Company, Ltd.
All rights reserved.

利用者検索インターフェースの特長

- キーワード検索と項目指定検索の 2種類を提供
 - 検索式による高度情報検索 (プロフェッショナル向け)
- フリーキーワード(全文検索)による柔軟な検索
 - 前方一致・完全一致の区別なく全て部分一致で検索
 - わかちがき入力やキーワード切り出し不要
 - 標題よみ、叢書名、原書名、内容細目を包括検索
 - 表記の違いを吸収する曖昧検索 (正規化)
 - 大文字/小文字、カタカナ/ひらがな、全角/半角、カタカナ表記、異体字、音標記号
- 対象資料を横断的に検索できる
 - 図書、雑誌、雑誌巻号、論文、ビデオ、CD-ROM
 - 検索レベル : 書誌情報、目次抄録、本文テキスト
- 著者、シリーズ、件名の各リンクを使った主題検索
- 外国文献などを対象として単語単位の検索も可能

画面-2)検索結果の一覧表示

検索結果一覧 (前)

UNEDD Search Result - Microsoft Internet Explorer

操作方法ヘルプ / 検索へ / トップページへ

検索結果一覧

検索条件:
KEYWORD=図書館

結果件数:
図書 93件 / 雑誌名 5件 / 雑誌 188件 / 論文 1121件 / 学内論文 32件 / ビデオ 44件 / CD-ROM 0件
全てブックマーク追加

ブックマーク追加

図書 1-10(93件)

次へ

1. [21世紀に向けての産官学連携戦略: ネットワーク社会における科学と産業](#) 奈良先端科学技術大学院大学AGIP21研究会編 化学工業日報社, 1998.
2. [ALA図書館情報学辞典](#) Heartsill Young [編]/丸山昭二郎, 高鷲忠美, 坂本博監訳 丸善, 1988.
3. [JAPAN/MARCマニュアル](#) 図書編 国立国会図書館編 国立国会図書館/日本図書館協会 (発売), 1988.
4. [Medical subject headings](#), 1992 和英版 日本医学図書館協会編 日本医学図書館協会, 1989.
5. [Medical subject headings](#), 1992 カテゴリー版 日本医学図書館協会編 日本医学図書館協会, 1989.
6. [Medical subject headings](#), 1992 英和版 日本医学図書館協会編 日本医学図書館協会, 1989.
7. [NAIST電子図書館レポート'97](#) 奈良先端科学技術大学院大学附属図書館, 1997.
8. [「NTT技術ジャーナル」にみる最新情報通信用語集](#) NTT研究開発推進部編 電気通信協会, 1996.
9. [医学情報へのアプローチ](#) 1992. 日本医学図書館協会, 1993.
10. [意義と方法](#) 補訂版 長沢雅男, 戸田慎一編 日本図書館協会, 1993. (図書館員選書16, 図書館学研究入門)

インターネット

検索結果一覧 (後)

UNEDD Search Result - Microsoft Internet Explorer

操作方法ヘルプ / 検索へ / トップページへ

1. [21世紀の図書館像を語る](#) 高忠雄 奈良先端科学技術大学院大学, 1995. (先端図書館国際シンポジウム: 21世紀の図書館像を語る)
2. [ELINDR Project](#) Anne Ramsden 奈良先端科学技術大学院大学, 1995. (先端図書館国際シンポジウム: 21世紀の図書館像を語る)
3. [Z39.50 JAPAN/MARC 対応関連仕様と評価試作システムの説明](#) NAIST, 1999. (第15回デジタル図書館ワークショップ)
4. [新しい文化の創造をめざして 望まれる多文化サービス](#) 紀伊國屋書店, 1998. (Library video series, 図書館の達人 司書実務編 pt.3,8)
5. [ある図書館長の一日: 市町村立図書館経営の実態](#) 紀伊國屋書店, 1998. (Library video series, 図書館の達人 司書実務編 pt.3,7)
6. [医学研究のための文献の探し方 \[映像資料\]](#) 紀伊國屋書店, 1994. (Library video series, 医学情報の達人2)
7. [医学文献データベース \[映像資料\] MEDLINE](#) 紀伊國屋書店, 1994. (Library video series, 医学情報の達人3)
8. [学生宿舎からの電子図書館利用](#) 奈良先端科学技術大学院大学, 1996.
9. [仮想の図書館で本を手にとって検索 / バーチャル・ライブラリー](#) 日本工業技術振興協会, 1997. (High-Tech Shower International:97041601)
10. [学会誌の電子出版](#) 寺田浩昭 奈良先端科学技術大学院大学, 1995. (先端図書館国際シンポジウム: 21世紀の図書館像を語る)

次へ

CD-ROM

検索条件を満たす資料はありません。

ブックマーク追加

のアイコンは内容へのリンクです, のアイコンは目次へのリンクです。

表示順形式設定:
[10] 件ずつ [タイトル] 順で 表示

全表示形式設定:
書誌 [簡略] 巻号 [GL] 所在 [あり] 設定 全表示

操作方法ヘルプ / 検索へ / トップページへ

インターネット

画面- 3) 検索結果の詳細表示

図書 (リンク検索)

雑誌 (受入巻号)

LMEDID Book Catalog Information - Microsoft Internet Explorer

ファイル 編集 表示 印刷 入力 ツール ヘルプ

[操作方法ヘルプ](#) / [検索](#) / [直前の検索結果](#) / [トップページ](#)

図書目録情報

書誌

- 書名 C++オブジェクト指向プログラミング
- 著者名 RS.ウイナー, LJ.ピンソン 著, 前川守 訳
- 出版 東京:トッパン, 1989.9
- 刊年 1989
- 形態 ix,291 p.; 21 cm
- 別書名 An introduction to object-oriented programming and C++
C++オブジェクト指向プログラミング
- シリーズ名 [アジソンウェスレイ・トッパン情報科学シリーズ](#)
- 標題言語 (jpn)
- 本文言語 (jpn)
- 原作言語 (eng)
- 著者情報 [Wiener, Richard, 1941-](#)
[Pinson, Lewis J.](#)
[前川 守\(1942-\) \(マエカワ, マモル\)](#)
- 分類 NDC8.007.64
NDC7.418.6
- 件名 [電子計算機—プログラミング](#)
[プログラミング\(電子計算機\)](#)
- 番号 NCID:BN0392551.4
ISBN:4810180093
JLA:89027060

所蔵

巻号	所在	請求記号	資料ID	貸出区分	状況(返却予定日)
1	資料室	DA110 23	0011715	一般	

[操作方法ヘルプ](#) / [検索](#) / [直前の検索結果](#) / [トップページ](#)

1ページが表示されました

LMEDID Magazine Catalog Information - Microsoft Internet Explorer

ファイル 編集 表示 印刷 入力 ツール ヘルプ

[操作方法ヘルプ](#) / [検索](#) / [直前の検索結果](#) / [トップページ](#)

雑誌目録情報

書誌

- 誌名 情報の科学と技術 (ジョウホウ ノ カガク ト キジュツ)
- 編集団体名 情報科学技術協会
- 創刊・終刊 Vol. 37, no. 1 (1987. 1)-
- 出版 東京: 情報科学技術協会, 1987-
- 並列誌名 The Journal of Information Science and Technology Association
- 標題言語 (jpn)
- 本文言語 (jpn)
- 団体名 [情報科学技術協会 \(ジョウホウ カガク キジュツ キョウカイ\)](#)
- 番号 NCID:AN10005857
ISSN:09133801/00125180

所蔵

所在	請求記号	所蔵年	所蔵巻号
1 図書館		1992-2000	42-49,50(1)+

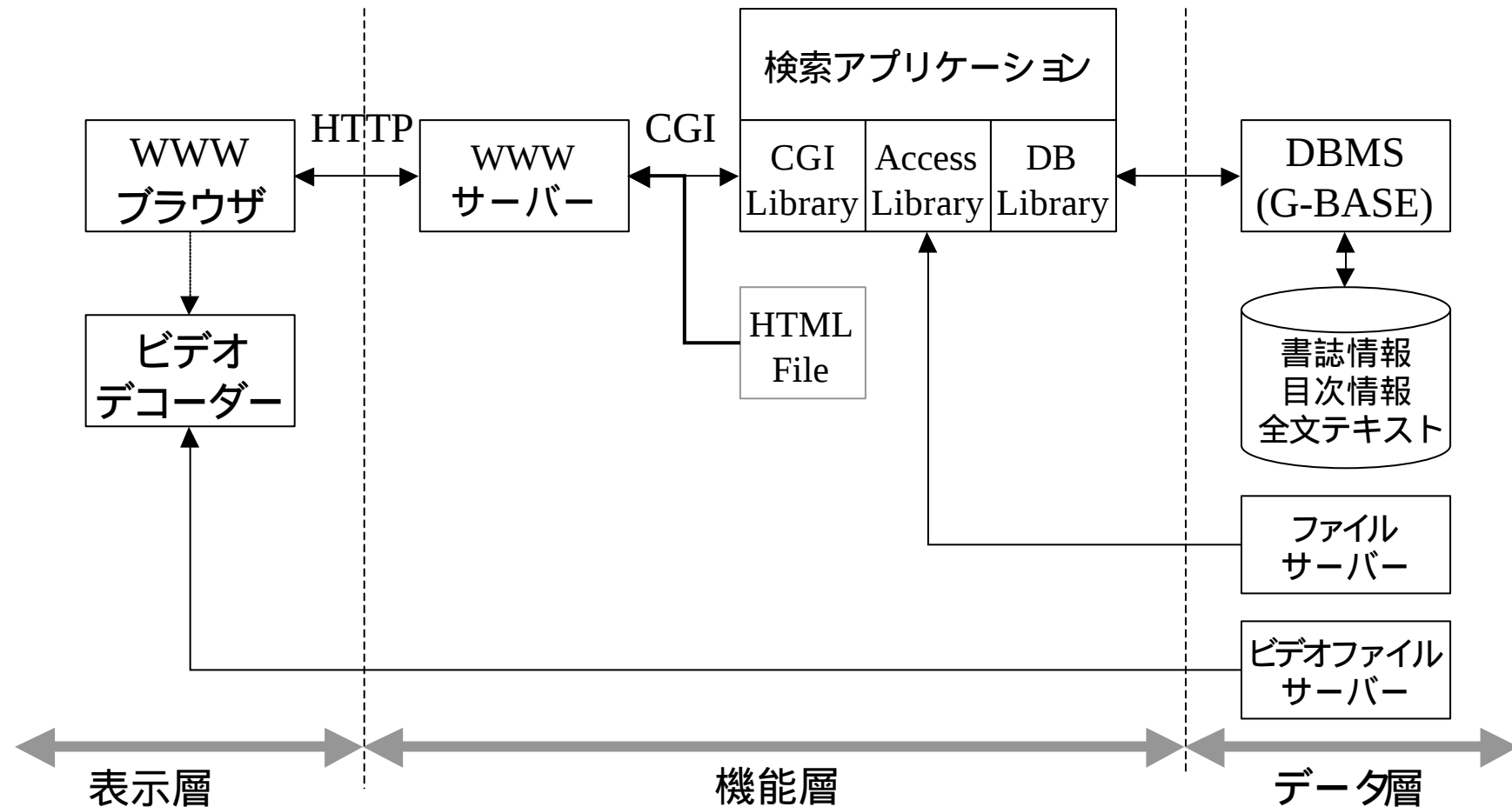
各号の情報

☐ 巻 ☐ 号 ☐ 年 ☐ 月の前後を 件 表示 [次へ](#)

	巻号	発行日	所在	資料ID	受入日	製本状態	状況(返却予定日)
1	Vol.49 No.12	1999年12月	図書館	3002715	1999/12/03		
2	Vol.49 No.11	1999年11月	図書館	3002666	1999/11/04		
3	Vol.49 No.10	1999年10月	図書館	3002615	1999/10/01		

1ページが表示されました

WWW 版OPACの基本アーキテクチャ



- 【長所】・3階層に分かれているため、負荷分散や機能拡張が容易
 ・ハイパーリンクによりネットワーク上の資源へのアクセスが容易
- 【課題】・サーバー側に「状態」を持たないため、ユーザーの情報保持が困難で、アプリケーション起動のためのオーバーヘッドが大きい

画面-4)電子図書館機能「目次情報」

目次情報表示

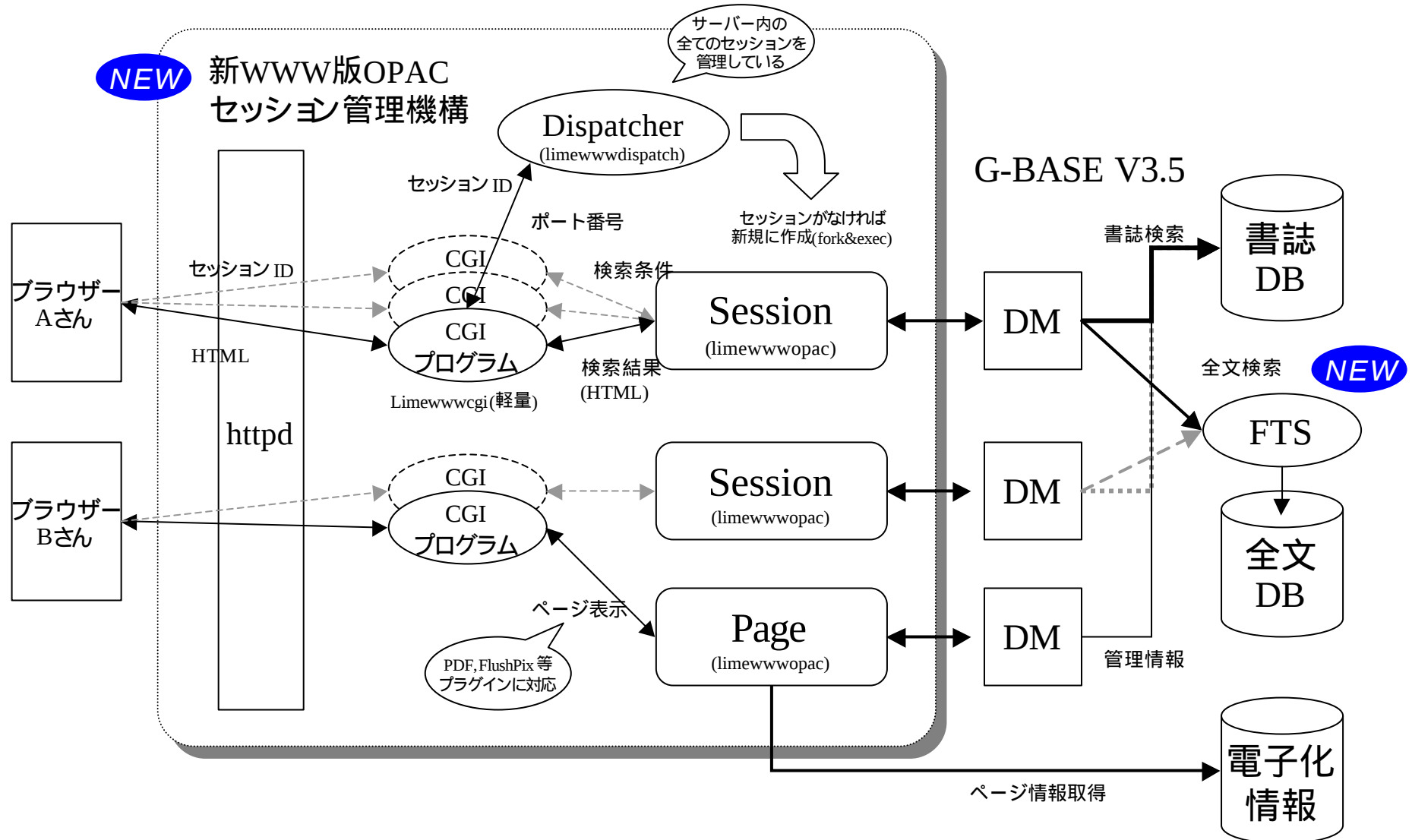
論文詳細表示



WWW版検索インターフェース

(電子図書館機能付WWW版OPAC)

■ Web上で仮想的なセッションを生成・管理している



セッション管理の導入で可能となった機能

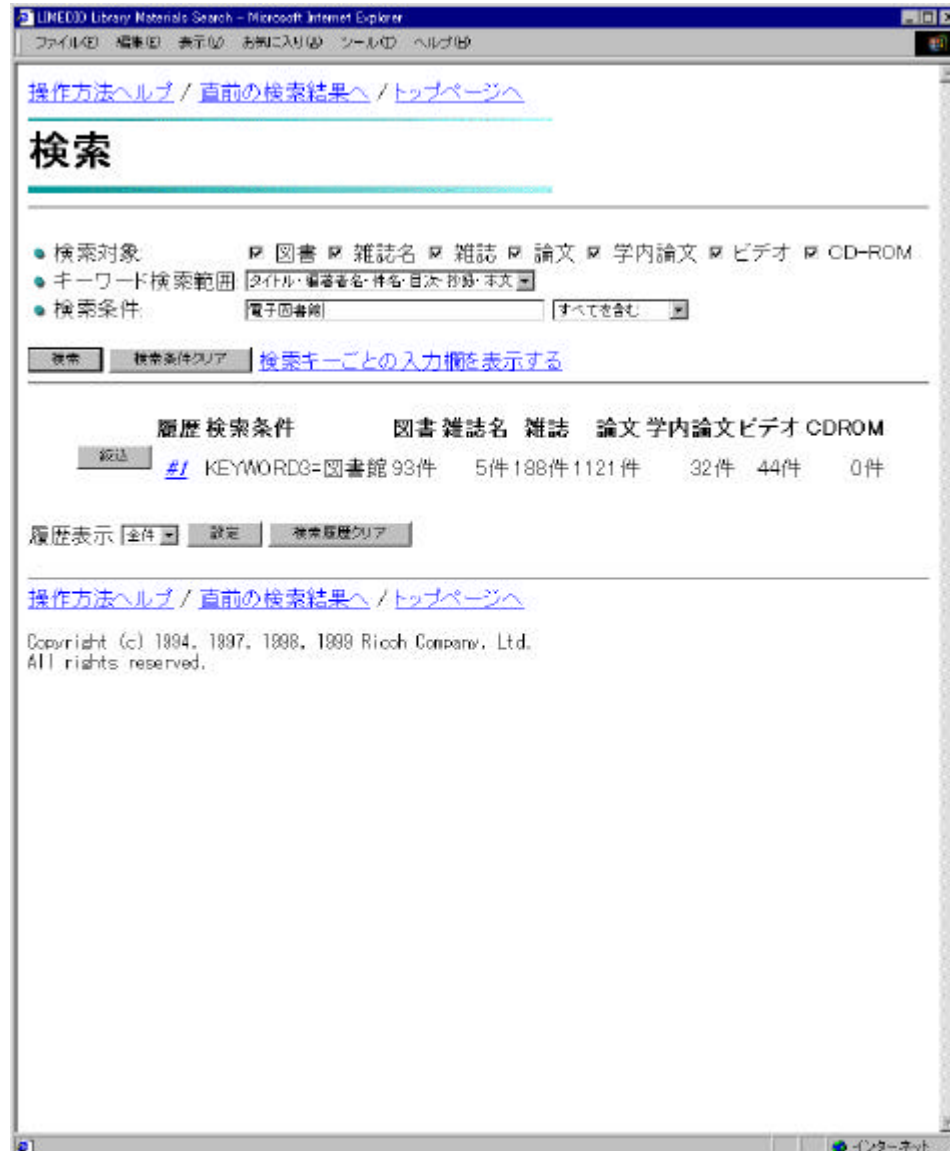
- 検索結果(= 履歴)を保持・参照
- 検索結果に対する絞込検索が可能
- 結果集合に対する集合演算が可能
- ブックマークを利用した結果保存が可能
- セッション単位のアクセス統計
- 検索性能の大幅向上
 - プロセス起動のオーバーヘッドを削減
- 分散アーキテクチャに対応可能

検索履歴・ブックマークは
セッション中のみで保持

画面-6) 検索履歴 絞込検索

検索履歴

絞込検索



WWW版OPAC機能のまとめ

集合演算 (and/or/not)	・No 検索導入により、ラフな検索後ノイズを除去 ・検索式インターフェース導入で、上級ユーザーにも対応 ・検索結果 (= 履歴) 間での集合演算が可能
全文検索機能	・英単語の区切りを意識した検索条件設定 (和文の場合、字種区切りで対応可能) ・転置ファイル+トライグラム導入による英文検索の高速化
近接演算 (near, followby)	・w1 near(N) w2: w1, w2 が N文字以内にある (順不同) ・w1 followedby w2: w2 が w1 の後に出現する ランキング検索検討中 ...
リンク検索	・著者名リンク、シリーズ名リンク、件名リンク ...
一覧表示機能	・結果一覧表示数を指定可能、巻号一覧の表示数指定可能 ・表示順を変更可能(タイトル、著者名、出版年)・ ・簡略一覧、詳細一覧指定可能、所在の表示・非表示 ...
カスタマイズ	・すべての項目を表示・非表示、項目名変更可能 ・コレクション一覧用のページ作成可能
その他	・PDF , FlushPix などプラグインへの対応 ・分館・所在など所蔵情報を意識した検索へ対応 ・ユーザーコード、ユーザー項目による検索

お手本は WWWの情報検索サービスサイト

画面-7)その他の機能

検索結果詳細一覧

検索条件式指定



UNEDD Search Result - Microsoft Internet Explorer

操作方法ヘルプ / 検索へ / トップページへ

検索結果一覧(全表示)

● 検索条件
KEYWORD3=図書館 AND KEYWORD3=電子図書館

● 結果件数
元の表示へ / 図書 9件 / 雑誌名 0件 / 雑誌 49件 / 論文 305件 / 学内論文 23件 / ビデオ 11件 / CD-ROM 0件 [全てブックマーク追加](#)

[ブックマーク追加](#)

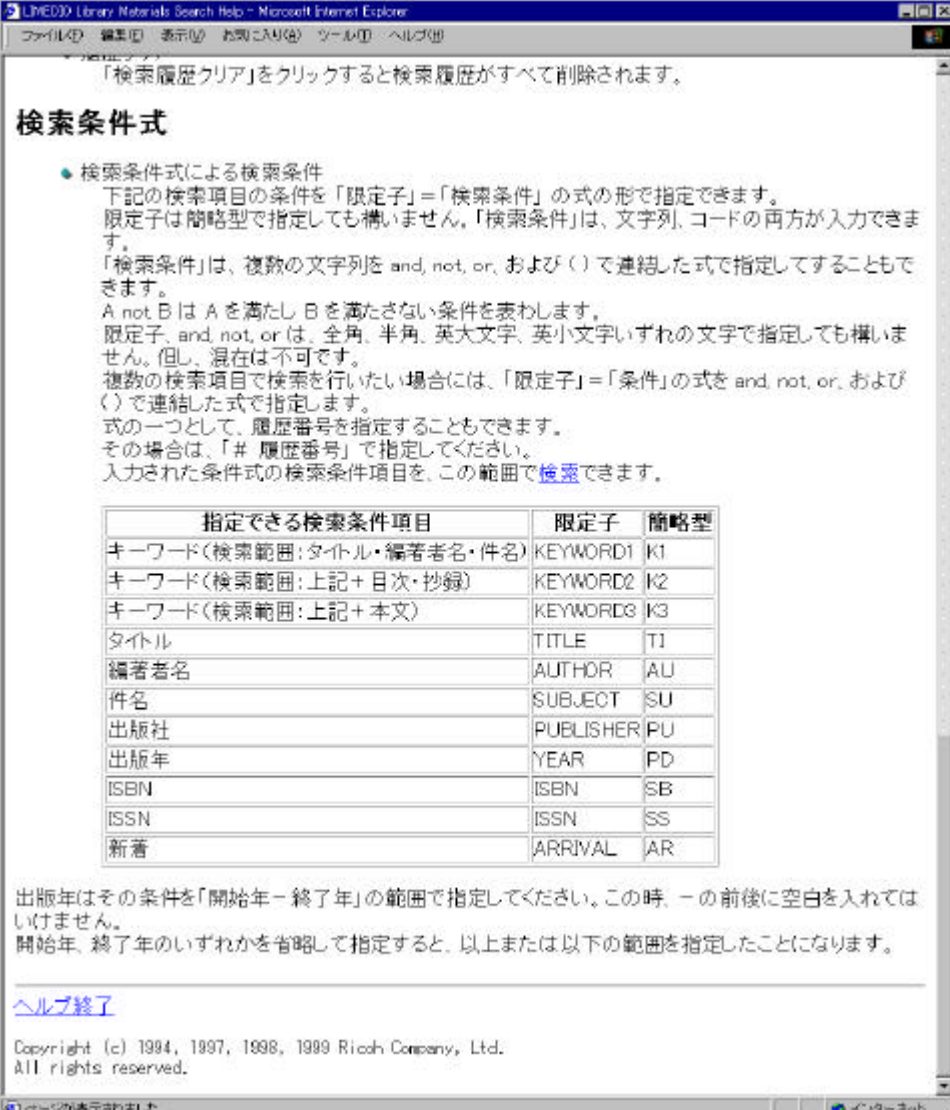
図書 1-9(9件)

1. [21世紀に向けての産官学連携戦略: ネットワーク社会における科学と産業](#) 奈良先端科学技術大学院大学AGIP21研究会編 化学工業日報社, 1998.

巻号	所在	請求記号	資料ID	貸出区分
1	図書館	377 NAR	0026258	一般
2. [NAIST電子図書館レポート'97](#) 奈良先端科学技術大学院大学附属図書館, 1997.

巻号	所在	請求記号	資料ID	貸出区分
1	図書館	013 NAR	0023478	一般
3. [「NTT技術ジャーナル」にみる最新情報通信用語集](#) NTT研究開発推進部編 電気通信協会, 1996.

巻号	所在	請求記号	資料ID	貸出区分
1	図書館	F547 NTT	0021698	一般
4. [情報管理入門: 情報技術による知的生産性向上](#) 佐藤修, 藤森洋志, 阿部昭博共著 弘学出版, 1997.



UNEDD Library Materials Search Help - Microsoft Internet Explorer

「検索履歴クリア」をクリックすると検索履歴がすべて削除されます。

検索条件式

- 検索条件式による検索条件
下記の検索項目の条件を「限定子」=「検索条件」の式の形で指定できます。限定子は簡略型で指定しても構いません。「検索条件」は、文字列、コードの両方が入力できます。「検索条件」は、複数の文字列を and, not, or, および () で連結した式で指定してすることもできます。
A not B は A を満たし B を満たさない条件を表わします。
限定子, and, not, or は、全角、半角、英大文字、英小文字いずれの文字で指定しても構いません。但し、混在は不可です。
複数の検索項目で検索を行いたい場合には、「限定子」=「条件」の式を and, not, or, および () で連結した式で指定します。
式の一つとして、履歴番号を指定することもできます。
その場合は、「# 履歴番号」で指定してください。
入力された条件式の検索条件項目を、この範囲で [検索](#) できます。

指定できる検索条件項目	限定子	簡略型
キーワード(検索範囲: タイトル・編著者名・件名)	KEYWORD1	K1
キーワード(検索範囲: 上記+目次・抄録)	KEYWORD2	K2
キーワード(検索範囲: 上記+本文)	KEYWORD3	K3
タイトル	TITLE	T1
編著者名	AUTHOR	AU
件名	SUBJECT	SU
出版社	PUBLISHER	PU
出版年	YEAR	PD
ISBN	ISBN	SB
ISSN	ISSN	SS
新着	ARRIVAL	AR

出版年はその条件を「開始年-終了年」の範囲で指定してください。この時、- の前後に空白を入れてはいけません。
開始年、終了年のいずれかを省略して指定すると、以上または以下の範囲を指定したことになります。

[ヘルプ終了](#)

Copyright (c) 1994, 1997, 1998, 1999 Ricoh Company, Ltd.
All rights reserved.

5 .全文検索システム

全文検索システム

- 膨大なテキストデータの集合から目的の情報を高速に検索する技術
 - 応用分野 : 新聞記事 DB、判例 DB、特許情報 DB、学术论文 DB、企業内文書 DB
- 欧米語の場合、単語単位の検索となる。日本語の場合、単語の切り出しが困難
- 分類 1 : 語句検索/文字列検索
 - 語句検索 : あらかじめ辞書に登録されている「索引語」を検索キーとし、索引語を含む文書を探す
 - 文字列検索 : 文字列を意味のないパターンとしてマッチする箇所を持つ該当文書を探す (自由文字列)
- 分類 2 : 完全一致/不完全一致
 - 完全一致 : 入力した検索文字列に完全に一致する文字列を含む文書のみを該当文書
 - 不完全一致 : 入力した検索文字列に一部を含む文書も該当文書とする全文検索
- 検索条件の指定
 - ブーリアン演算 (ある/なし)、類似検索 (近い/遠い)、近接演算 (何語以内/前後関係)、ランキング

全文検索技術による分類

名称	検索方式の概要	採用製品の例
文字成分表	入力されたテキストの文字に対して 1文字または隣接 2文字からなるテーブル (文字成分表) を作成。この文字成分表の検索により該当文書を抽出。	Bibliotheca/TS (日立) 読取物語キャビネット(リコー) 俊司くん(言語工学研究所)
N文字インデックス (N-グラム)	入力されたテキストをN文字ごとに区切り、各文字列が含まれる該当する文書の情報と文字列の位置情報によるインデックスを作成する。	SearchServer (フルクラム) InfoSearch(IBM) Oracle7 ConText Option(オラクル) G-BASE/ETS (リコー)
パトリシアツリー	文書のすべての文字から始まる文字列 (これを semi-infinite string と呼ぶ) を作成する。	OpenText (日商岩井インフォコム)
パターン認識	入力されたテキストの文字列の情報をパターンとして認識したインデックスを作成する。件作事には、検索語のパターンを計算し、インデックスから抽出する。	NSSEARCH (新日本製鉄) SAVVY/EFS (日経情報システム)
形態素解析	日本語を解析するための辞書を使用して、入力されたテキストを単語に分解し、このなかから名詞などのキーワードを抽出してインデックスを作成する。	Future/Happiness (平和情報センター) SEARCH'97(オムロン) FAIRS-SV (富士通)
概念検索	形態素解析で単語を認識すると共に、統計処理で複合語の認識を実施。入力文書の語彙空間を抽出して検索対象の文書中から類似性の高い文書を検索する。	ConceptBase (ジャストシステム) VextSearch(コマソソフト) J-LiveSearch(日本総合研究所)
その他	(非公開)	PanaSearch/SS (松下電器産業) DOCS Open (NKエクス) Documentum(CTCラボラトリ) MyQuick(帝人システムテクノロジー)

(出典 : 全文検索システム協議会)

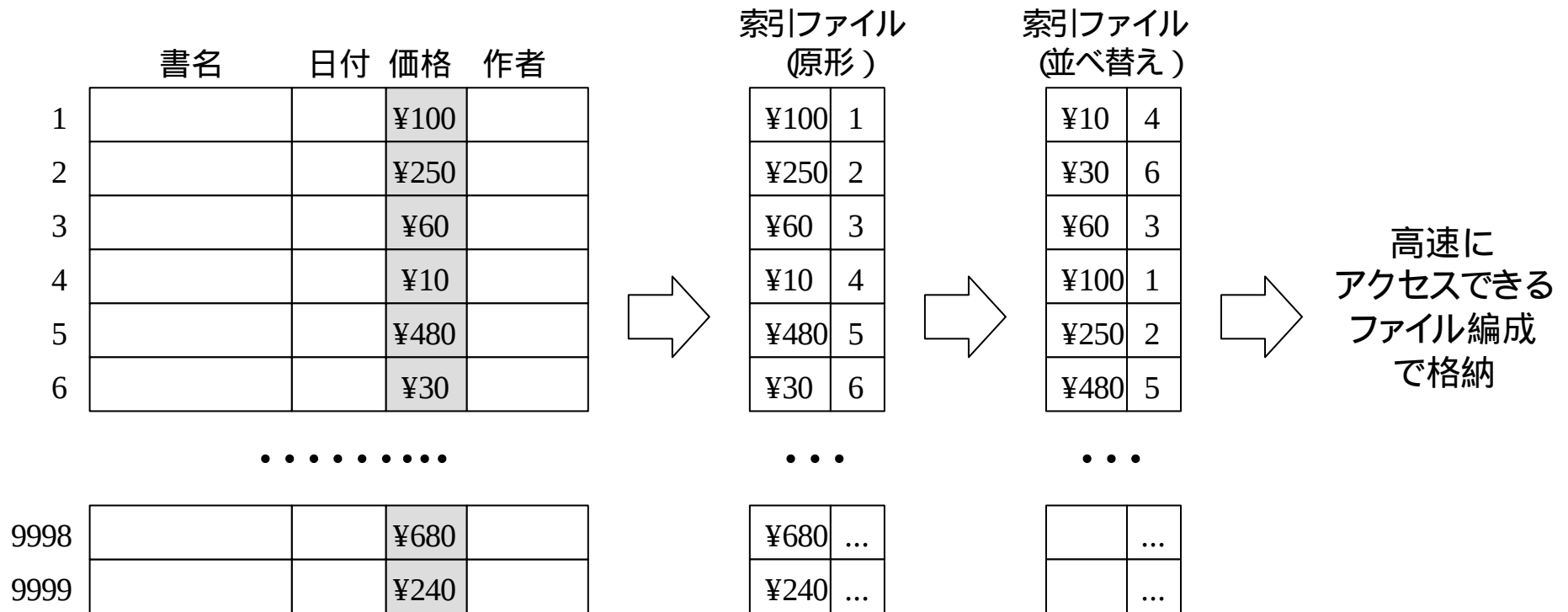
G-BASE の全文検索機能

- RDBMS の索引 (Index)として全文検索エンジンを組み込んでいる。
 - 文字列項目 (最大 2GB) に対して索引定義ができる
 - 他の項目との関係演算による問合せを行える。
 - 索引へのデータ追加・削除・修正が共有モードで行える
 - ✦ 夜間処理のように大量データを処理する場合には排他処理の方が早くなる
- N-グラム方式で「シグネチャー索引」と「転置ファイル索引」の2つの索引を選択可能
 - 英文は 3-Gram, 和文は 2-Gram
 - 書誌情報など短い文字列の場合には「シグネチャー索引」
 - 全文テキストなど長い文書の場合には「転置ファイル索引」
- 転置ファイル索引の場合、隣接演算やランキング検索にも対応。

✦ 隣接演算は実装済み、ランキングは計画中

一般的なデータベースの索引ファイル

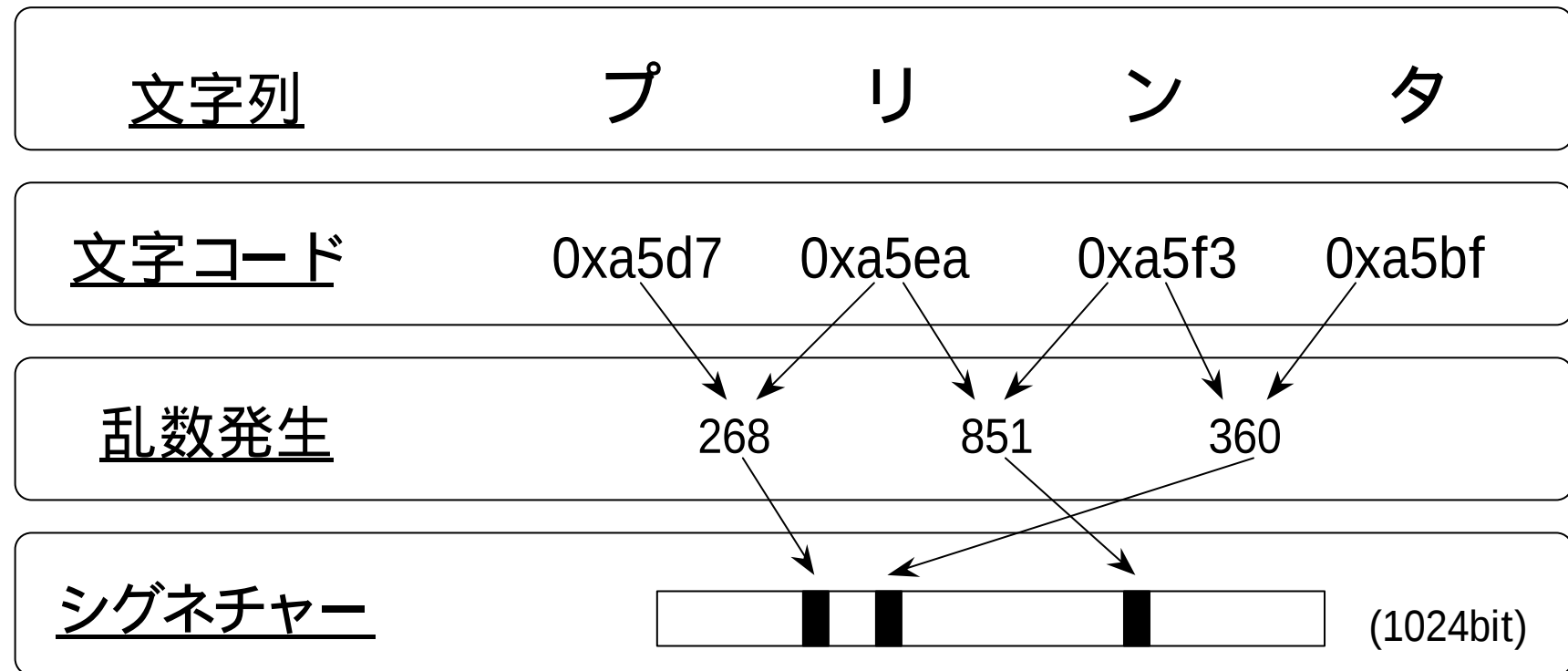
- データ実体は表形式で管理される (行 = レコード、項目 = フィールド)
- 項目の値から目的の行を高速に得る (逆引きする) ために索引ファイル利用
- 数値データや短文字コードなどの場合、B木方式が一般的に利用される



項目順の索引では
完全一致・前方一致
しかできない

文字成分表 (2グラム) によるシグネチャー索引

隣接2文字をすべて単語と考慮してシグネチャー発生
+ スーパーインポーズを行なう。

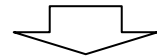


G-BASE V3.5 (FTS版) : 「シグネチャー索引」vs. 「転置索引」

シグネチャー索引

DBに格納する文字列と検索文字列を同じアルゴリズムで同じ長さのビットパターンに変換し、ビット単位で検索する

「プリンタ」



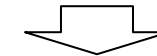
パターンに変換

	0	0	1	0	0	1	0	1	0	0	0	0
	Search-1			Search-2			Search-3					
REC1	0	0	1	1	0	1	1	0	1	0	1	1
REC2	1	0	0	1	0	0	1	0	1	1	0	0
REC3	0	0	1	1	1	1	0	1	0	0	0	0
REC4	0	1	0	0	0	1	1	0	1	1	1	0
REC5	0	1	0	0	1	0	1	0	1	0	1	0
REC6	1	0	0	1	0	1	0	0	1	0	1	1
REC7	0	1	1	1	0	0	0	1	1	1	0	0
REC8	1	1	0	0	0	0	0	1	0	0	1	1
REC9	0	0	0	1	1	1	1	0	1	1	0	1
REC11	0	1	0	0	0	1	0	1	0	1	1	0
REC12	1	1	0	1	0	0	1	1	1	0	0	0

転置索引

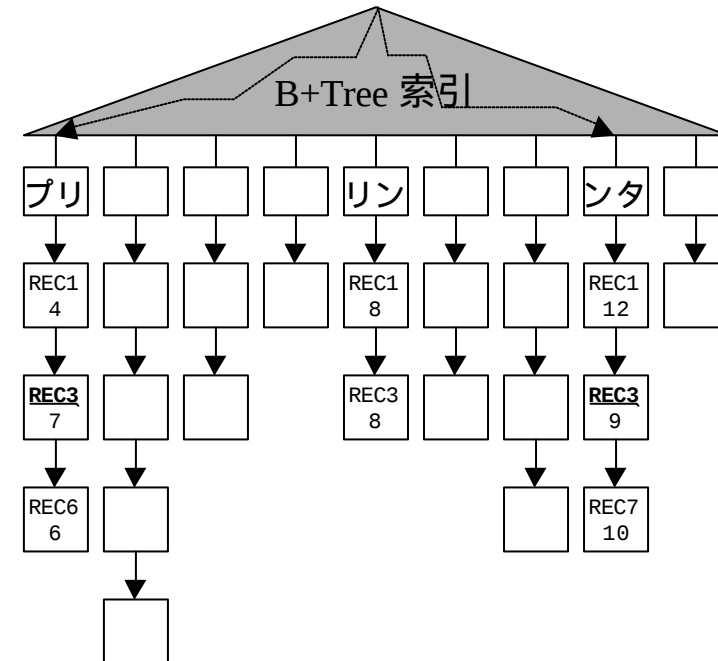
DBに格納する文字列と検索文字列をN文字づつ (通常 2 文字) 分割し、出現位置とレコードIDを要素にした高速インデックスファイル (B木) などで管理する

「プリンタ」



B+Tree索引はデータ数に関係なく、ほぼ一定時間に検索結果が得られる索引ファイル (ISAM)

「プリ」 & 「ンタ」で検索し距離2のものを求める

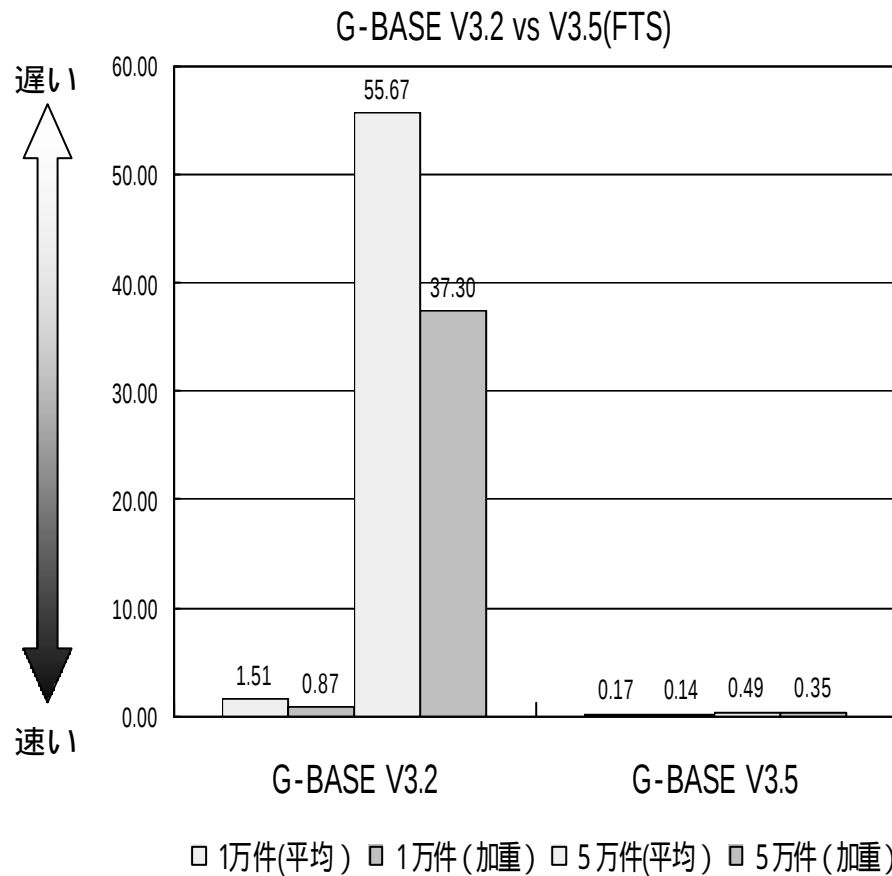


欠点：

1. ノイズが発生するため検索後フィルターをかける必要がある
→ 検索結果が多い場合には遅くなる
2. ビットマップ中はリニアサーチになるため対象データ数に影響される

1. データの文字数が多いと変更作業に時間がかかる
(追加、変更、削除)

G-BASE V3.5 (FTS版)のベンチマーク結果



	1万件(平均)	1万件(加重)	5万件(平均)	5万件(加重)
G-BASE V3.2	1.51	0.87	55.67	37.30
G-BASE V3.5	0.17	0.14	0.49	0.35
/	8.9	6.2	113.6	106.6

■ 対象データ:特許全文明細データ(96年)

- 1万件 (約0.2GB)
- 5万件 (約 1GB)

■ 検索条件:480種類の検索文字列使用

- 検索結果が0にならない一般的な用語
- 単一検索 320, 複合検索 160種
- 漢字、カタカナ、異字種混合

■ 検索結果

- 平均:検索にかかる時間の単純平均
- 加重:検索条件の利用頻度による加重平均
例)漢字2文字による検索は、漢字1文字による検索より利用頻度が多い。

■ ベンチマーク環境

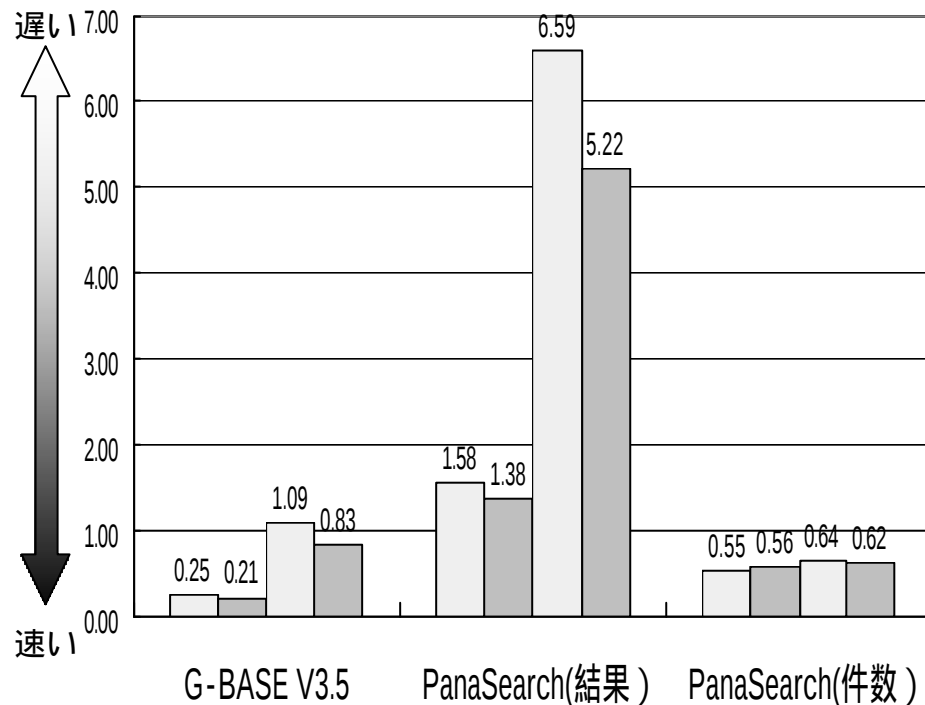
- Sun Ultra 10 (UltraSPARC II 300MHz)
- Solaris 2.5.1
- メモリ128MB, HDD 4.3GB x4 (Software RAID0に設定)

■ インデックス

- G-BASE V3.2 はシグネチャーインデックス
→ 短い文字列には強いが、長文DBの場合、データ量に比例して時間がかかる。
- G-BASE V3.5(FTS)は転置ファイルインデックス
→ データ量が増えても、検索時間があまりかからない (Diskアクセス分は増加するが ...)

G-BASE V3.5 (FTS版)とPanaSearch の比較

G-BASE V3.5 (FTS) vs. PanaSearch /SS



□ 1万件(平均) ■ 1万件(加重) □ 5万件(平均) ■ 5万件(加重)

	1万件(平均)	1万件(加重)	5万件(平均)	5万件(加重)
G-BASE V3.5	0.25	0.21	1.09	0.83
PanaSearch(結果)	1.58	1.38	6.59	5.22
PanaSearch(件数)	0.55	0.56	0.64	0.62
/	2.2	2.7	0.6	0.7

■ 対象データ:特許全文明細データ(96年)

- 1万件(約0.2GB)
- 5万件(約1GB)

■ 検索条件:480種類の検索文字列使用

- 検索結果が0にならない一般的な用語
- 単一検索 320, 複合検索 160種
- 漢字、カタカナ、異字種混合

■ 検索結果

- 平均:検索にかかる時間の単純平均
- 加重:検索条件の利用頻度による加重平均
例)漢字2文字による検索は、漢字1文字による検索より利用頻度が多い。

■ ベンチマーク環境

- MICRON ClientPro (PentiumPro 200MHz)
- WindowsNT 4.0 Server (SP3)
- メモリ512MB
- NEWTECH NRAID126G/128 RAID5設定 (Ultra Wide SCSI3)

■ その他

- PanaSearch/SS (松下電器産業)NT版
- Yahoo!Japan や朝日新聞記事検索サービス
- PanaSearch (件数)は、ヒット件数のみを返し検索結果を得ていない。
- 異表記正規化処理 OFF
- 複合検索(And/Or/Not)が遅い

6 .LIMEDIO電子図書館オプション

LIMEDIO電子図書館オプション

「電子図書館システム」として標準的に利用される
(簡易な)機能をパッケージのオプションとして提供

■ 「目次情報機能」

- 図書や雑誌の目次の内容をテキストデータとして登録し、利用者による検索や参照を可能にする
- 目次登録のための対話型インターフェースを提供、利用者インターフェースは WWW 版 OPAC と統合
- 目次情報は学術情報センターの雑誌目次速報 (紀要) システムの形式 or LIMEDIO形式から一括登録できる

■ 「ページイメージ情報機能」

- 図書や雑誌のページの見たい目 (ページイメージと呼ぶ) を画像ファイルとして登録し利用者による参照を可能にする機能
- WWW版の「ページイメージ登録ツール」を提供
 - 冊子体の資料をスキャナー等でデジタル画像ファイルに変換する機能は含まれていない。
- 目次情報機能と両方を導入することで、目次と本文のページイメージを関連つけて管理できる

電子図書館システム 「今後の課題」

■ 一次情報作成支援

→ ページイメージ・全文テキストデータ

■ 著作権保護機能の強化

→ アクセスレベルの詳細化 例) 検索対象、閲覧、印刷、ダウンロード...

→ 電子データの保護 例) ウォーターマーク、電子透かし...

→ 記事や内容の著作権 (?)

■ 様々なデータ形式に対応

→ 電子文書データ 音声・動画データ

■ 他システムとの連携

→ 他の情報検索システム (オンラインDBやCD-ROMサーバ等)

→ 他の電子図書館システム

■ WWW上のデータの探索

→ インターネットからの情報収集・蓄積・提供

■ ブラウジング(接架)・関連検索 ...



ご静聴、ありがとうございました。

