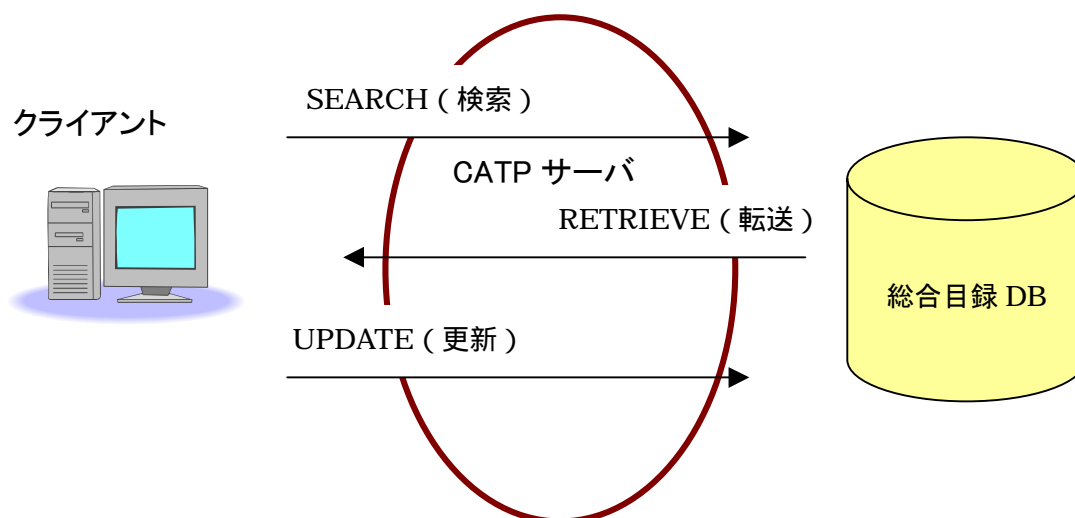


新 CAT/ILL システム及び多言語対応システム解説

1. CATP 解説

1.1 基本メソッド概念図



1.2 旧システムとのコマンド-メソッド比較

機能	旧 CAT コマンド	新 CAT メソッド	CATP 定義表
接続	LOGIN LOGOFF	GETHANDLE RELEASEHANDLE	
検索 (インデクス検索)	SEARCH	SEARCH	検索定義
検索 (二次検索)	SCAN	SCAN	
表示 (簡略)	BROWSE	RETRIEVE	転送定義
表示 (詳細)	DISPLAY	RETRIEVE	
新規作成	CREATE COPY	INSERT	更新定義
修正	EDIT	UPDATE	
削除	WEED	DELETE	

1.3 フィールドの新設

新設したフィールドは、2 種類ある。(データベースフィールド定義)

(1) その他のヨミフィールド

多言語対応にともない、その他のヨミフィールドを新設した。

各データベースのタイトル関係の各フィールドグループに、その他のヨミフィールドを追加し、中国語のピンイン等を入力できるようにした。

また、中位の書誌のその他のヨミは、PTBNO に格納することとした。

(2) IDENT フィールド

電子ジャーナルの URL など、アクセス方法を記録するために、IDENT フィールド (Identifier: 資源識別子) を新設した。

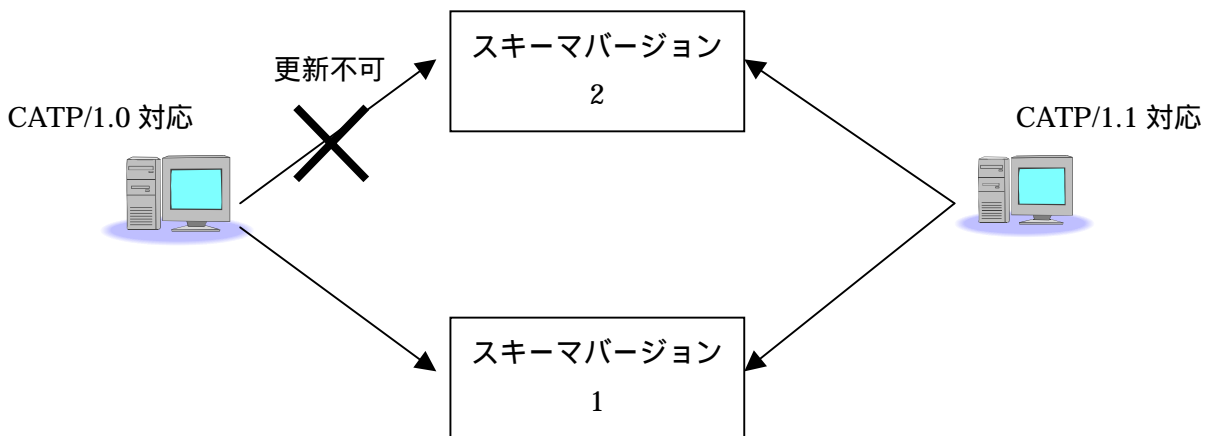
IDENT フィールドの使用方法については、「総合目録データベースにおける電子ジャーナルの取扱い」を参照のこと。

1.4 CATP スキーマバージョン

DB のスキーマ変更（フィールド新設）が行なわれた場合に，その変更に対応していないクライアントにより DB へのアクセスが行なわれると，フィールドデータが削除される等の不都合が生じる可能性がある。

この問題を解決するために CATP/1.1 では DB スキーマの実装レベル（以下，DB 実装レベル）を表わすスキーマバージョンを導入している。レコードのスキーマバージョンは，各フィールドのスキーマバージョン値の最大値となる。

旧 CAT システムでも同様だが，IDENT フィールドに限って例外的に，表示・編集とも可能となっている。



項番	メソッド名	リクエスト受信時	レスポンス返送時
1	GETHANDLE	クライアントのスキーマバージョン取得	なし
2	RELEASEHANDLE	クライアントのスキーマバージョン破棄	なし
3	RELEASEFRAME	なし	なし
4	SEARCH	なし	返送フィールドチェック
5	SCAN	なし	返送フィールドチェック
6	RETRIEVE	なし	返送フィールドチェック
7	INDEXLIST	なし	なし
8	INSERT	レコード登録チェック	返送フィールドチェック
9	UPDATE	レコード登録チェック	返送フィールドチェック
10	DELETE	レコード登録チェック	なし
11	SERVERPROCEDURECALL (ILL)	レコード登録チェック	返送フィールドチェック

(1) レコード登録チェック

レコードを新規・更新登録または削除する際に，対象レコード中にクライアントが実装するスキーマバージョンを超えるフィールドが存在するかどうかチェックし，もし存在すれば実行不可としてエラーレスポンスを返す。

(2) 返送フィールドチェック

レコードをクライアントへ返送する際に，クライアントの DB 実装レベルを超えるフィールドが存在するかどうかチェックし，もし存在すれば返送レコードにはそのフィールドを含めない。

スキーマバージョン値定義

項番	DB 種類	CATP フィールド名	スキーマバージョン
1	図書書誌	TRVR	2
2		VTVR	2
3		CWVR	2
4		PTBTRVR	2
5		AHDNGVR	2
6		UTHDNGVR	2
7		SHVR	2
8		IDENT	2
9		上記以外	1
10	雑誌書誌	TRVR	2
11		VTVR	2
12		AHDNGVR	2
13		SHVR	2
14		IDENT	2
15		上記以外	1
16	著者名典拠	HDNGVR	2
17		SFVR	2
18		SAFVR	2
19		上記以外	1
20	統一書名	HDNGVR	2
21		SFVR	2
22		SAFVR	2
23		上記以外	1
24	上記以外	全フィールド	1

2. 多言語対応

2.1 UCS 対応

多言語対応サーバ内部で使用する文字コードを、EUC から UCS に変更する。

このサーバで使用する UCS は、以下の仕様とする。

- (1) 従来の EXC 文字を含む
- (2) UCS の実装水準は 3 とする。これは、EXC の全てを表現するのに、合成文字を使用するためである。
- (3) JIS X0208 の漢字包摂基準を採用する。

クライアントが使用する文字コードとサーバ内部の文字コード（UCS）との変換は、CATP エンコーディングの指定により多言語対応サーバで行う。この UCS 変換は双方向の変換であり、CATP リクエスト時と CATP レスポンス時に実行される。

クライアント文字コードに変換できない文字を、UCS 外字という。これは、 UCS コード値をはさんだ形に変換する。

2.2 文字コード変換処理概要

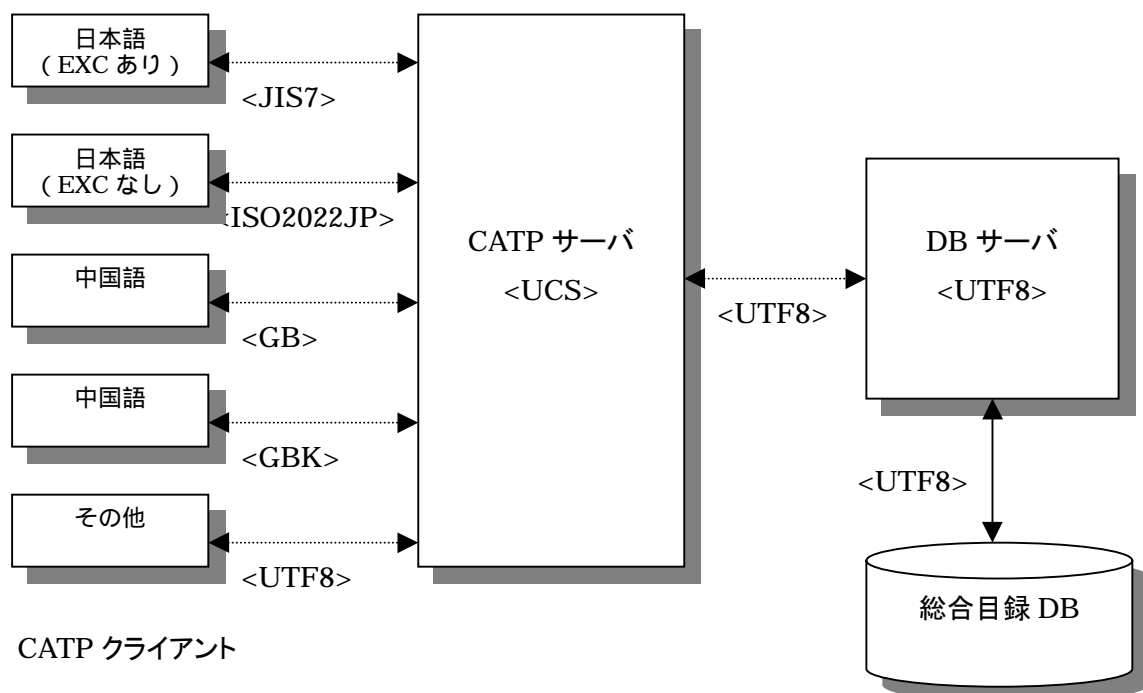
従来のサーバでは、JIS7 エンコーディングだけに対応していたが、新たに GB、GBK、UTF8、ISO2022JP の 4 種類のエンコーディングに対応可能となった。

ISO2022JP とは、EXC 文字を含まない JIS7 文字セットを扱うものであり、EXC 文字は、UCS の番号（UCS 外字）で表示されることになる。

GB とは、簡体字中心の中国語用文字セットであり、日本語の表示はできない。

GBK はその拡張版で、UCS と同等の CJK（中日韓）統合漢字を含んでいるものである。

UTF8 とは、UCS の文字セットを扱うエンコーディングである。



[関連資料]

資料 1. EXC 文字/UCS コード対応表

資料 2. 拡張ラテン文字定義表(EXC 文字以外)

資料 3. UCS 外字仕様

2.3 正規化等処理

項番	処理	概要	備考
1	文字コード変換	サーバ内部文字コードと外部文字コードの相互変換を行う。	
2	UCS 包摂	CATP クライアントから送られた外部文字データを正規化する。	
3	検索キー正規化	検索キーに対して行う正規化処理である。	

2.4 UCS 包摂処理(文字の正規化処理)

UCS 包摂*とは、データベースに格納するデータ、および CATP サーバの内部処理のために行う正規化である。

サーバでの CATP メッセージの受信時に CATP オブジェクトボディのデータに対して、UCS コードへの変換、UCS 包摂の順に行う。UCS 包摂で行う処理の一覧を次に示す。

* 包摂：「ある概念をより一般的な概念の中にとり入れること。」(新明解国語辞典)

UCS 包摂処理一覧

処理順	項目	対象	概要	備考
1	ラテン文字包摂	ラテン文字	ラテン文字の半角化を行う。	
2	数字包摂	数字	数字の半角化を行う。	
3	記号包摂	記号	記号の半角/全角化を行う。	
4	漢字包摂	漢字	JIS X0208 の漢字包摂基準に従って漢字を包摂する。	→「基準」付録 1
5	カタカナ包摂	カタカナ	カタカナの全角化を行う	
6	合成文字包摂	合成文字	合成文字の正規化(カノニカルオーダリング)を行う。**	

** UCS では、EXC のような合成文字は、1 文字で表したり複数文字の組み合わせで表したりすることができる。それらを統一するために、異なる組み合わせパターンの合成文字(基底文字+結合文字)を、一つの合成文字に正規化する。

例) 「Á」 「A」 + 「´」

2.5 検索キー正規化

検索キー正規化とは、CATP リクエストのオブジェクトボディに含まれるレコード (INSERT, UPDATE, SERVERPROCEDURECALL メソッド)、検索条件 (SEARCH, INDEXLIST) から検索キーを生成する際に行う正規化である。検索キー正規化で行う処理の一覧を次に示す。

検索キー正規化処理一覧

項番	項目	対象	概要	備考
1	大文字化	ラテン文字	ラテン文字の小文字を大文字へ変換する。	
2		ギリシャ文字	ギリシャ文字の小文字を大文字へ変換する。	
3		キリル文字	キリル文字の小文字を大文字へ変換する。	
4		カタカナ	拗音, 促音を大文字へ変換する。	
5	小文字化	ラテン文字	ラテン文字の大文字を小文字へ変換する。	コード類のみ
6	カタカナ化	ひらがな	ひらがなをカタカナへ変換する。	
7	ラテン文字化	EXC	EXC をラテン文字へ変換する。	
8	漢字統合	漢字	似た形や同じ意味の漢字をそのグループを代表する漢字へ変換する。	→ 漢字統合インデクス

3. 漢字統合インデクスの実装

UCS の漢字統合部分には、約 2 万字の漢字が含まれており、似た形や同じ意味の漢字が数多くある。このことによる検索漏れを防ぐため、似た形や同じ意味の漢字を含めて統合検索を可能とするのが、漢字統合インデクスである。

漢字統合インデクスの作成対象となるフィールドは、以下の通りである。

漢字統合対象フィールド

項番	DB 種類	DB	フィールド名
1	図書書誌	BOOK RECON JPMARC TRCMARC USMARC USMARCX GPOMARC UKMARC DNMARC CHMARC	AKEY TITLEKEY AUTHKEY PUBLKEY PUBPKKEY SHKEY FTITLEKEY
2	雑誌書誌	SERIAL JPMARCS USMARCS	同上
3	著者名典拠	NAME JPMARCA USMARCA	AKEY AUTHKEY PLACEKEY TIMEKEY
4	統一書名典拠	TITLE USMARCT	AKEY TITLEKEY AUTHKEY

[関連資料]

資料 4. 漢字統合インデクス