

委託業務の題目

「ライフサイエンス統合データベース開発運用」

(統合データベース開発：専門用語辞書管理システムと専門用語解析技術の開発)

受託機関名：奈良先端科学技術大学院大学

1. 委託事業の9月末時点の判断基準になる目標（以前にお出しいただいたもの）

- ・専門用語辞書システムを、属性情報等の追加に柔軟に対応するため、MongoDBでの実装を完成させる。
- ・専門用語の内部構造解析済みデータを2500語以上に拡大する。
- ・専門用語辞書システムとシソーラスの階層構造表示システムを統合する。

2. 9月末時点の達成状況

- ・専門用語辞書システムの属性情報追加機能は設計（仕様）を完了。MongoDBへの実装は進行中。
- ・専門用語の内部構造解析済みデータの蓄積：9月末時点で2656語。（10月28日現在で、3454語）
- ・専門用語辞書システムとシソーラスの階層構造表示システムの統合：シソーラス階層表示システムから専門用語辞書システム(Cradle-LSD)へのリンクは完成。

3. 上記達成状況を踏まえたプロジェクト終了までの目標

- ・専門用語辞書システムのMongoDBへの移植はプロジェクト終了までには達成。
- ・専門用語内部構造解析済みデータは、今後も若干作業を予定しており、少なくとも4000語規模に拡大する予定。
- ・専門用語辞書システムからシソーラス階層構造表示システムへのリンクを実装し、両システムの統合を完成させる。

4. 成果の概要

ユーザー評価サイトで、以下の2種類3つのシステムを公開した。公開日はいずれも10月25日。

・ライフサイエンス辞書検索システム：Cradle-LSD

ライフサイエンス辞書を対象に、用語そのものだけでなく、部分文字列、品詞、シソーラスコードなど様々な情報によって用語検索を行う。また、用語の内部構造の表示や内部構造アノテーションを行う機能をもつ。

・バイオ医療専門用語と類似文脈の検索

・PNE と LSD に対するバイオ医療専門用語の検索

指定した用語のPNE(雑誌「蛋白質・核酸・酵素」)内での出現を検索し、前後文脈とともに表示する。また、その用語のMeSHシソーラス内での位置を表示する。各シソーラスコードには対応する専門用語(日本語)が表記され、Cradle-LSDにリンクされている。

・PNE に対する類似文脈の検索

専門用語とその出現文脈を入力すると、類似した文脈をPNEより検索し、それによって用語と類似の専門用語を検索するシステム