

文献や特許データベース中の化学構造式の認識と検索

育成研究：JSTイノベーションプラザ福岡 平成20年度採択課題
「文献や特許データベース中の化学構造式の認識と検索」

代表研究者：九州大学 大学院数理学研究院 特任教授

鈴木 昌和



■ 研究概要

光学的文字認識技術(OCR)と構造解析技術により、印刷物中の化学構造式画像を認識し検索可能な業界標準形式に変換する技術の研究を行った。これにより、従来は手作業で行っていた化学構造式画像の電子化作業が効率化する。電子化されたデータは業界標準形式のため、既存のツールと連携することができる(図1)。

■ 研究内容、研究成果

我々が開発した ChemInfty(ケムインフティ)システムは、論文や特許などの印刷物中の化学構造式画像を認識するソフトウェアとその認識結果を効率的に修正するための修正インターフェースソフトウェア(図2)から構成されている。作成した化学構造式画像の認識および修正ソフトウェアは、競合ソフトウェアと比べ、(1)接触画像などの低品質画像への対応(図3)や(2)操作性の高い修正インターフェース、(3)特許に頻繁に用いられるマーカッシュ記法の認識(図4)に対応しているという点で優位性がある。実際に2008年の特許のデータを電子化し、既存のシステムへ入力し検索システムのプロトタイプを作成を行った(図5)。また、化学構造式を手書きで簡単に入力できるインターフェースのプロトタイプを作成した(図6)。このインターフェースは、PCのウェブブラウザやiPadやAndroidなどのスマートフォンでも操作することができる。

プロジェクトのウェブページ：<http://www.inftyproject.org/jp/ChemInfty/>

■ 今後の展開、将来の展望

本プロジェクトでは、以下の事業展開を考えており事業パートナーを探している段階である。

認識および修正ソフトウェア販売

プロジェクト内で開発した認識および修正ソフトウェアの年間契約による利用権を販売する。

ネット経由の認識サービス

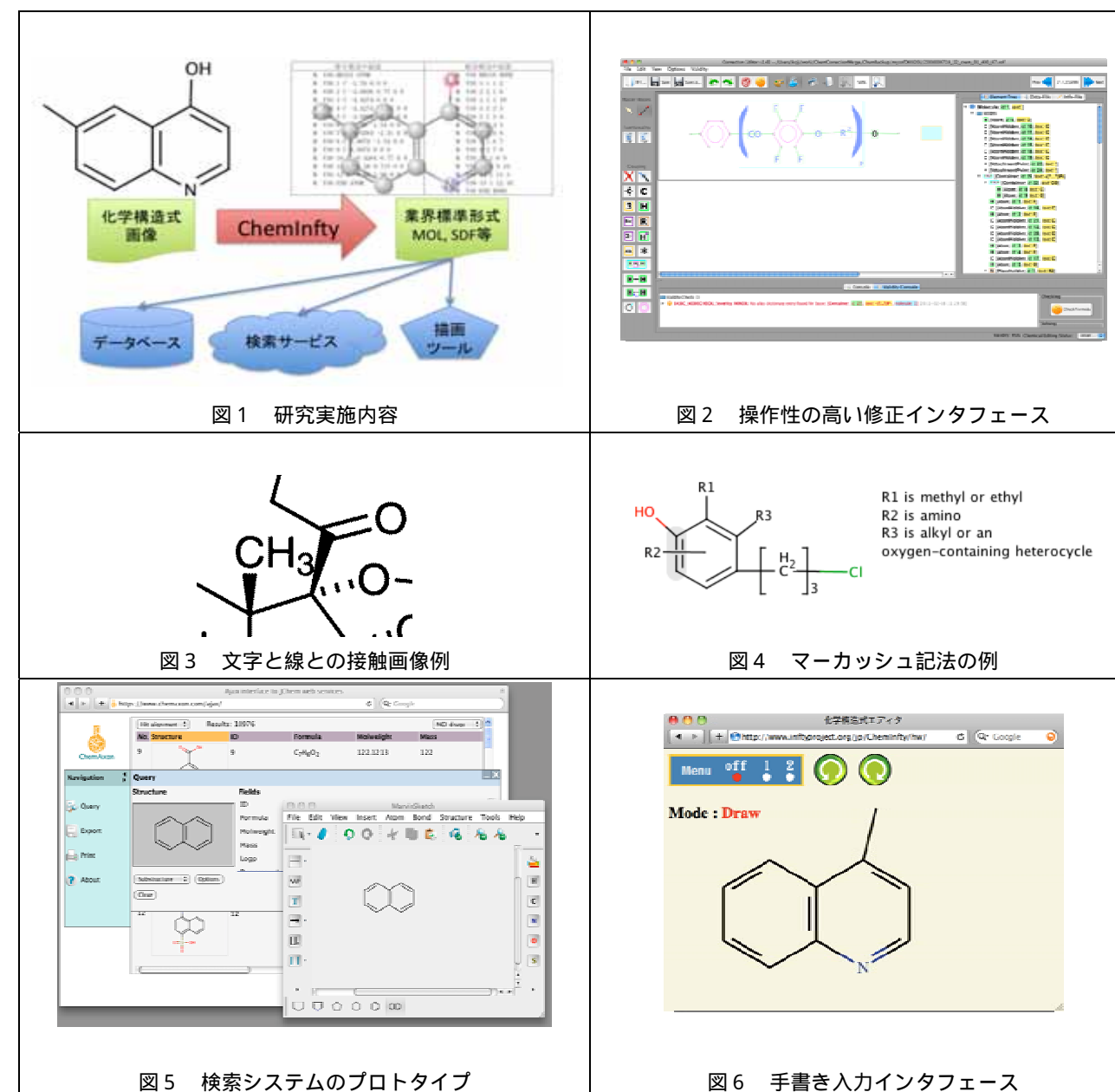
ユーザがネットを通じて画像をアップロードし、その認識結果をネットからダウンロードする認識サービスである。

電子化サービス(認識と修正)

顧客の持つ電子化されていない化学構造式の画像データを電子化し、その電子化データを納品するサービスである。

電子化システム構築サービス

化学構造式を電子化するシステム全体をカスタマイズした上で納品する。



■ 研究体制

◆ 代表研究者

九州大学・大学院数理学研究院 特任教授 鈴木昌和

◆ 研究者

中川 康二(九州大学・大学院数理学研究院) 藤芳 明生(茨城大学・工学部)

坂本 好夫、白木 智文(財団法人九州先端科学技術研究所)

伊東 弘子、高橋 竜也(株式会社システムオーディット)

濱田 謙二(株式会社デジタルノーツ) 宮島 真一(株式会社キューデンインフォコム)

◆ 共同研究機関

茨城大学、財団法人九州先端科学技術研究所、株式会社システムオーディット、株式会社デジタルノーツ、株式会社キューデンインフォコム

■ 研究期間

平成21年4月 ~ 平成24年3月

【連絡先】〒819-0395 福岡市西区元岡 744
九州大学 数理学研究院
Tel : (092) 802-4402 Fax : (092) 802-4405
e-mail : contact@infty.biz