

2011 年度 日本文化デジタル・ヒューマニティーズ拠点 研究プロジェクト 研究計画書

2011 年 4 月 21 日提出

1. 研究プロジェクト名		歴史的文書に関わるデジタル図書館研究
2. 研究プロジェクト代表者		前田 亮
3. 研究班 メインとなる研究班 その他		京都文化研究班
		日本文化研究班
		歴史地理情報研究班
		デジタルアーカイブ技術研究班
		Web活用技術研究班
4. 研究期間		2011年 4月 ~ 2012年 3月
5. 研究メンバー		
種別	氏名	所属・職名
事業推進担当者	前田亮	立命館大学大学院理工学研究科・教授
特別招聘教員		
研究員		
客員研究員		
PD		
RA	ビルゲサイ ハン・バドジ ヤルカル(Bi ligsaikhan Batjargal)	立命館大学大学院理工学研究科・D2
学内研究協力者	木村文則 井坪将 大崎隆比古 村岡流 久木貴博	立命館大学情報理工学部・助手 立命館大学大学院理工学研究科・M2 立命館大学大学院理工学研究科・M2 立命館大学大学院理工学研究科・M1 立命館大学大学院理工学研究科・M1
その他	ハルタルフ ー・ガルマー バザル(Khal tarkhuu Gar maabazar) 手塚太郎	モンゴル日本人材開発センター・総括主任 (Senior Manager, Mongo lia-Japan Center for Human Resources Development, Mongolia)  筑波大学図書館情報メディア研究科・准教授

6. 2011年度教育研究計画（今年度の教育研究内容、目的と結果の予想の関係が理解できるようにご記入ください。特に若手研究者（研究メンバーのPD、博士課程後期課程大学院生）の役割、教育効果を具体的にご説明ください）。

2011年度は、主に以下の3つのテーマについて、それぞれ研究を進める。また、今年度が最終年度であるため、開発した各システムについて専門家などによる利用者評価実験を行い、また可能な限りシステムをインターネット上で公開する。

### 1. 日本語古典史料のテキスト処理手法の開発

当プロジェクトでこれまで行ってきた日本語古典史料の現代語による検索、テキスト分析および可視化の研究を進展させ、日本語古典史料に対するテキスト処理手法の確立を目指す。

当プロジェクトでこれまでに開発してきた現代語・古語辞書の自動構築手法を進展させ、言語資源に乏しい日本語古典史料に適用可能なテキスト処理技術を開発する。具体的には、文字Nグラムの出現確率を用いた日本語古典史料テキストからの単語抽出の手法について、昨年度に引き続き検討を行い、実用化を目指す。この技術を基に、現代語による古典史料へのアクセス、古典史料からの固有表現抽出、テキスト分析、可視化の技術を開発し、これらの人文学研究への応用について検討を行う。また、古典史料のテキスト分析による人物関係の抽出と可視化の研究について、Web上でインタラクティブに利用可能なシステムを構築し、人文系研究者による利用者評価を行う。

本研究は、学内研究協力者の木村文則助手およびM2の井坪将・大崎隆比古を中心として研究を進める。本研究により、日本語古典史料のテキスト処理手法の発展への貢献が期待される。

### 2. 人文系データベースの横断検索システム

本学アート・リサーチセンターをはじめ、国内外で多数公開されている人文系データベースの相互利用を図るため、これらのデータベースを横断的に検索するための技術について研究を行う。

当プロジェクトでこれまでに開発してきた、浮世絵の多言語・データベース横断検索システムを進展させ、既存の日本の浮世絵データベースとしてアート・リサーチセンターの浮世絵検索閲覧システムおよび早稲田大学演劇博物館の浮世絵閲覧システム、海外の浮世絵データベースとして大英博物館およびボストン美術館の各収蔵品データベースなどを対象として、英語と日本語による言語横断検索およびデータベース間の横断検索を実現する。さらに、Linked Dataの技術を用いることで、異なるデータベース間のリンクによる「芋蔓式検索」をシステムに実装する。

本研究は、RA 種のビルゲサイハン・バドジャルカルおよび学内研究協力者の木村文則助手、M1の村岡流・久木貴博を中心として研究を進める。本研究により、デジタルヒューマニティーズにおける重要な技術であるデータベース構築とメタデータ設計に関するスキルの向上が期待される。

### 3. 伝統的モンゴル文字文書のデジタル図書館システム

従来から研究を進めている伝統的モンゴル文字文書のデジタル図書館システムについて、コンテンツの拡充および大規模な利用者評価実験を行う。

また、従来の検索機能に加えて、伝統的モンゴル文字文書における固有表現抽出およびテキスト分析技術の開発を行う。これらを基に、主に欧米において人文学研究に広く用いられているテキストのマークアップ規格である TEI (Text Encoding Initiative) 標準に準拠することで、テキストの再利用や他のデータベースとの連携など新たな応用への発展を目指す。また、今後はモンゴル国内の研究者との連携を進め、伝統的モンゴル文字のテキスト処理技術の確立と人文学研究への応用について検討を行う。

本研究は、RA 種のビルゲサイハン・バドジャルカルおよびモンゴル日本人材開発センターのハルタルフー・ガルマーバザルを中心として研究を進める。本研究により、デジタルヒューマニティーズ研究において特にテキスト分析の研究で広く使用されているTEI (Text Encoding Initiative) 標準に関する理解とスキルの向上が期待される。

7. 教育研究計画・方法		
教育研究目的を達成するための計画・方法、実施する場所をできるだけ具体的に記入してください		
実施時期	計画内容	実施場所
2011年4月～9月	Nグラム出現確率を用いた日本語古典史料テキストからの単語抽出手法の実装	BKC
2011年4月～9月	人文系データベースの横断検索手法の検討およびプロトタイプシステムの実装	BKC
2011年4月～9月	伝統的モンゴル文字文書のデジタル図書館システムのコンテンツ拡充およびテキスト分析手法の検討・評価実験	BKCおよびモンゴル・ウランバートル
2011年6月	国際会議DH2011において、日本語古典史料のテキスト分析および可視化に関する研究発表	アメリカ・スタンフォード
2011年10月～2012年3月	日本語古典史料のテキスト処理手法の実際の古典史料への適用とテキスト分析および可視化の実験	BKC
2011年9月	国際会議DC-2011において、人文系データベースの横断検索に関する研究発表（予定）	オランダ・ハーグ
2011年10月～2012年3月	伝統的モンゴル文字文書のデジタル図書館システムのTEI (Text Encoding Initiative) 準拠手法の検討	BKCおよびイギリス・SOAS
2011年10月～2012年3月	人文系データベースの横断検索システムのシステム改良および評価実験	BKCおよびイギリス・UCL
2011年10月～2012年3月	Nグラム出現確率を用いた日本語古典史料テキストからの単語抽出手法の評価実験	BKC
2011年10月	国際会議Culture and Computing 2011において、日本語古典史料のテキスト処理手法および人文系データベースの横断検索に関する研究発表（予定）	京都大学
2011年12月	人文科学とコンピュータシンポジウムにおいて、日本語古典史料のテキスト処理手法および人文系データベースの横断検索に関する研究発表（予定）	未定
2011年12月～2012年2月	人文系データベース横断検索システムの海外の日本文化研究者による評価実験	イギリス・SOAS/UCL
2012年3月	データ工学と情報マネジメントに関するフォーラムにおいて、人文系データベースの言語横断検索に関する研究発表（予定）	未定