

2004年度東アジア人文情報学サマーセミナー報告書

1 目的

21世紀 COE「東アジア世界の人文情報学研究教育拠点」プロジェクトでは、漢字文化の全き継承と発展のために、インターネット時代における人文学に情報学的手法を導入することで、漢字文化圏に関する人文情報学を新たに構築することを目的としている。したがって、その教育プログラムとしては、従来の文献学的方法論の習得とともに、新たに情報学の技能を身につけた人材の育成を目指さなければならない。

そこで、人文学と情報学を横断する教育の場を設けるために、人文科学研究所と人間・環境学研究科の人材育成プログラムの事業推進者を中心として、東アジア人文情報学の高度専門カリキュラムの開発を試み、研究科の態勢が整い次第、正規のカリキュラムのなかで人文情報学の授業を立ち上げる方向で協議しているところである。また、そのパイロット・プロジェクトを兼ねて、夏期休暇期間を利用したサマーセミナーを開催し、博士後期課程の学生を対象とする集中講義を企画することにした。

サマーセミナーの狙いとしては、COE プロジェクトの研究が東アジア言語のコンピュータ処理、東アジア文献のナリッジベース化を二つの柱とするのであるから、その技法の習得を計ろうとするものである。しかし、1週間という短い期間であり、受講生が必ずしもコンピュータ科学に精通しているわけではないから、多言語処理やデータベースのための諸技法のなかで、最も基礎的であると思われるものに対象を絞り、パソコン実習を行うことにした。

対象となる大学院生の予備調査では、パソコンを Word、一太郎等のワープロソフトで使用している者が大半であり、研究論文や資料を汎用性のあるデータベースにする工夫を行っている者はごく少数に限られていた。しかし、ホームページ作成には、大いに関心があり、すでに立ち上げている学生も少なからず存在した。したがって、ウェブページを書くための HTML (HyperText Markup Language) について、もっと学びたいという希望が最も強かった。

そこで、本年度のサマーセミナーのテーマとして「インターネット時代の人文学の技術」を掲げ、TeX と XML という二つの手法を取り上げることにした。

TeX とは、スタンフォード大学の Donald E. Knuth 教授が開発した文書組版ソフトであり、TeX の機能を強化した LaTeX、日本語が使えるようにした pTeX 等がさらに考案されている。これらは、自分の思ったとおりに印刷用の出力が機種、OS の環境に依存しないで得られるフリーソフトであるために、数式、化学式等を多用する理系研究者を中心に広く用いられている。

COE プロジェクトが構想するデータベースという見地において、TeX の文書整形技法

で注目されるのは、タグづけによって文章の構造や修飾情報を記述していくマークアップ言語であるという側面である。SGML(Standard Generalized Markup Language)から発展したHTML(Hyper Text Markup Language)もマークアップ言語の一つであるが、その言語によって記述された文書はテキストファイル形式であるため、文系の研究者にとっても、汎用性の高い論文、資料を作成することのできる利点がある。

一方、XML(eXtensible Markup Language)は、SGMLをインターネットで活用しやすくするために1998年2月に開発されたものであるが、タグが固定的で表示に特化したHTMLの制約を克服しており、XML文書は汎用性のあるデータ交換の標準フォーマットとして現在最も期待されている。

以上のような現状を踏まえて、セミナーでは、TeX、LaTeXの文書整形技術の訓練を導入と、TeXからXMLへという流れにおいて、文章の論理構造、意味構造を記述するマークアップ手法に理解を深め、XMLの基礎を習得するとともに、プログラム言語のPerl、XMLのHTML変換用プログラムXSLTの初歩を実習することにした。また、漢字に限定することなく、多言語処理の基礎と方法を幅広く概観するために、講師陣には、TeX演習ではイスラム史研究者の真下裕之氏、インド・チベット仏教研究者の宮崎泉氏に加わってもらった。また、XML演習では、高度なプログラミング技法を活用した漢字情報学のデータベースを作成し、活発な研究教育活動を行っている師茂樹氏、岩井茂樹氏を招き、具体的な応用例を示しながらプログラミングの初歩的な手ほどきをしてもらうことにした。

なお、パソコンによる実習指導を徹底するために、受講生は10名程度とした。選考方法は、人間・環境学研究科の指導教員に依頼して博士課程学生を推薦してもらうとともに、公募による受講希望者のなかから、選考委員会において書類審査によって選定した。また、坂内繁夫岐阜大学教育学部助教授がセミナーの参加を申請したので、選考委員会で協議し、聴講を許可した。

2 日程及びカリキュラム

- ・開催日 2004年9月6日(月)～10日(金)
- ・会場 人間・環境学研究科棟433・434演習室
- ・テーマ インターネット時代の人文学の技術
- ・カリキュラム

9月6日(月)10:30～16:30

開会式

第1講 漢字情報学序説「入力、出力、そして検索」 安岡孝一

第2講 TeX 入門「人文系研究者のための文書整形術」 真下裕之

実習1 TeX 演習

9月7日(火)10:30～16:30

第3講 LaTeX 応用「効率的な文書作成のために」 守岡知彦

第4講 多言語化概論「 、TeX と OTF パッケージ」 宮崎泉

実習2 LaTeX 演習

9月8日(水)10:30～16:30

第5講 マークアップ概論「マークアップ概念の導入 LaTeX を例に」 守岡知彦

第6講 XML 入門「The Basics -- Understanding and Using XML」 C.ウイッテルン

実習3 XML・マークアップ演習

9月9日(木)10:30～16:30

第7講 XML 応用「Applications of XML -- XML for Text Encoding」 C.ウイッテルン

第8講 XML プログラミング入門「正規表現とXML文書の操作」 師茂樹

実習4 プログラミング演習

9月10日(金)10:30～16:30

第9講 XSLT 入門「XMLでXMLを処理する？」 岩井茂樹

第10講 多言語情報処理概論「漢字政策と漢字情報処理」 安岡孝一

修了式

3 参加者及びスタッフ

< 受講生 >

陳 捷 京都大学人間・環境学研究科D2 阿辻研

甲骨文字における殷代信仰の研究(修論 漢字における對稱思想)

鍾 獅 京都大学人間・環境学研究科D4・人文科学研究所非常勤研究員 金坂研

東南中国における歴史人文地域の形成とその構造に関する歴史地理学的研究

中国江南の北江盆地を中心に

中国南方における方言地理学的研究

桂 由起 京都大学人間・環境学研究科 D 2 愛宕研

中国の民衆版画（年画 / 美術史・図像学）

崔 水晶 京都大学人間環境学研究科 D 4 松浦研

十七、十八世紀日韓関係史（朝鮮通信使が見た対馬 - 十七、十八世紀を中心に -

閻 淑珍 京都大学人間・環境学研究科 D 1 西脇研

神仙道教

鄭 宰相 京都大学文学研究科 中国哲学史専攻 D 2

中国古代思想史（荀子を中心として）

高 永聡 京都大学文学研究科 中国文学専攻 D 2

蘇軾及びその周辺

石野 一晴 京都大学文学研究科 東洋史学専攻 D 2

明清社会・宗教史（修論題目「明清の江南における観音聖地への参詣活動」）

大井 和也 関西大学大学院 文学研究科哲学専攻 D 3

仏教学 中国中世仏教思想（特に華嚴教学と如来蔵思想）

秋山 陽一郎 立命館大学文学研究科研究生・COE 非常勤職員

中国古代文献学（博論「劉向本戦国策の文献学的研究」）

< 聴講者 >

坂内 栄夫 岐阜大教育学部助教授 中国道教思想史研究

< 講師 >

岩井 茂樹（人文科学研究所教授）

安岡 孝一（漢字情報研究センター助教授）

C. ウィッテルン（漢字情報研究センター助教授）

宮崎 泉（京都大学文学部専任講師）

師 茂樹（花園大学文学部専任講師）

真下 裕之（人文科学研究所助手）

守岡 知彦（漢字情報研究センター助手）

< スタッフ >

高田 時雄（研究リーダー 人文科学研究所教授）

武田 時昌（教育部門リーダー、漢字情報研究センター教授）

阿辻 哲次（人間・環境学研究科教授）

愛宕 元（人間・環境学研究科教授）

金坂 清則（人間・環境学研究科教授）

石川 禎浩（人文科学研究所助教授）

山本 一登（COE 研究員）

4 講義概要

第1講 漢字情報学序説「入力、出力、そして検索」 安岡 孝一

文字に対する情報処理の基本は、入力・出力・検索である。漢字に対しても、この点
は変わらない。しかし、たかか 26 種類の文字か、あるいはその倍も処理すればよい
欧米諸語とは異なり、漢字情報処理は 10,000 種類を超える文字を相手にしなければ
ならない。そこにはおのずから、漢字情報処理ならではの手法、工夫、面白味あるい
は限界がある。この講義では、漢字情報学の入門編として、その手法や面白味をわか
りやすく解説する。

第2講 TeX 入門「人文系研究者のための文書整形術」 真下 裕之

文書の体裁を整えるために、なぜワープロではなく TeX を用いるのか。この講義で
は、TeX の基礎知識を学ぶことによって、ワープロに対する TeX の数多くの長所（お
よびいくつかの短所）を解説する。また文書整形の他に、TeX が備えている便利な機
能のいくつかも紹介する。そのうえで、TeX の使用が人文系研究者にもたらす二次的
な利点についても考えたい。

第3講 LaTeX 応用「効率的な文書作成のために」 守岡 知彦

LaTeX による文書作成において、良く使われる幾つかの環境やコマンドを紹介する。
特に、論文などの大きな文書を対象にファイル分割や相互参照、文献引用の方法など
も紹介し、BibTeX による参考文献情報の管理法も紹介する。

第4講 多言語化概論「 、TeX と OTF パッケージ」 宮崎 泉

文字空間を 16bit に拡張して多言語化に正面から取り組んだ は、OTF という枠組み
を用意することで柔軟な拡張性を持つ。この OTF を利用して、CHISE の多漢字環境
を印刷する方法を説明するとともに、漢字以外の文字も含めた多スクリプト環境の印
刷の可能性を紹介する。また、多漢字環境だけを必要とする人のために、pTeX と OTF
パッケージを利用した、より日本語の組版に適した方法も紹介したい。

第5講 マークアップ概論「マークアップ概念の導入 LaTeX を例に」 守岡 知彦

LaTeX による文書作成を例にマークアップ概念を説明する。LaTeX 文書の使い回し
を例に文書の論理構造と見掛けの情報の分離する手法を説明する。また、LaTeX の
問題点を概説するとともに、より一般的なマークアップ言語の必要性を説明する。

第6講 XML 入門「The Basics Understanding and Using XML」 C . ウイッテルン

In recent years, XML has become the major standard for text encoding and text interchange.
Any researcher in the humanities, who wants to use his computer for more than as a fancy
typewriter, should add knowledge of XML to his toolbox; this will enable him to query,
analyze and interpret texts he wants to work with. At the same time, XML can also be used
to write articles and other texts. In this course, we will first introduce the foundations of
XML, the underlying concept of texts and other related technologies. We will then proceed to

introduce some of the possibilities and applications XML has been put to so far.

第7講 XML 応用「Applications of XML XML for Text Encoding」 C. ウィッテルン

In building on the basic knowledge of XML, we will now introduce some of the applications of XML in more detail, especially the markup vocabulary and recommendations published by the Text Encoding Initiative (TEI). Using some concrete examples, we will introduce basic techniques of document analysis and how to translate the results of such an analysis to markup and subsequently apply them to the text. If time allows, we will also demonstrate how to transfer the markup thus created for various purposes.

第8講 XML プログラミング入門「正規表現とXML文書の操作」 師 茂樹

プログラミング言語を用いて、XML でマークアップされた漢字文献を操作する基本的な方法を学ぶ。文字列の柔軟な検索・置換が可能な正規表現について漢字文献の特性を踏まえつつ紹介した後、XML 文書の構造を操作するための代表的な方法である DOM や SAX などについて概観する。余力があれば多言語処理などについても一瞥したい。

第9講 XSLT 入門「XML で XML を処理する？」 岩井 茂樹

XSLT とは、XML 文書にたいして、構造を変更したり、抽出・検索を行ったりするために設計されたスクリプト言語である。たとえば、XML 文書になっている電子テキストや資料を、Web サイト上で利用するには、XML 文書を HTML 文書や XHTML 文書に変換する必要が生じる。その際、まず章や節の表題だけを取りだして、本文へのリンク付き目次を用意できると便利である。また、閲覧者から送られた語句によって検索をおこない、結果を表示する仕組みも必要かもしれない。XML 文書にたいしてこうした操作をおこなうさいに、XSLT で書かれたスタイルシートが威力を発揮する。この XSLT で書かれたスタイルシート自体もじつは XML 文書であり、すべての命令や関数にタグが付いている。データであろうと、プログラムのソースであろうと、すべての文書を XML 化してしまおうという志向が吉とでるか凶とでるか、まずは試してみよう。

第10講 多言語情報処理概論「漢字政策と漢字情報処理」 安岡 孝一

戦後の近代化の波は各国の漢字政策にまで波及し、漢字文化圏は中国・台湾・日本・韓国の4つに分断されたまま、それぞれが独自の漢字を発展させる結果となった。ところが20世紀末のインターネットの普及は、これら4つの「国」における漢字の違いを、あらためて問い直す契機となった。そのような中、漢字情報処理はどのように揺れ動き、またその将来はどうなっていくのか。この講義では、複数の言語にわたる漢字情報処理の過去と未来について論説する。