

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



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。そのさい、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。

修了式

【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）

発行日 2004年7月26日

発行者 京都大学21世紀 COE プログラム

「東アジア世界の人文情報学研究教育拠点—漢字文化の全き継承と発展のために—」

住 所 〒606-8265 京都市左京区北白川東小倉町47 京都大学人文科学研究所

電 話 075-753-6997 FAX 075-753-6999

e-mail coe@zinbun.kyoto-u.ac.jp Web Site <http://coe21.zinbun.kyoto-u.ac.jp/>





# 2004年度 東アジア人文情報学サマーセミナー



```

<body>
  <div org="uniform" sample="complete" part="N">
    <pb n="5771-"/>
    <head>資治通鑑卷第一百八十五</head>
    <p><note place="inline">端明殿學士兼翰林侍讀學士太
  <lb n="5771-025"/>中大夫提學西京嵩山崇福宮上柱國河 郡開國公食邑二千
  <lb n="5771-050"/>二百戶食實封九百戶賜紫金魚袋臣</note>司馬光<note place="inli
  <p><note place="inline">後學天
  <lb n="5771-075"/>台</note>胡三省<note place="inline">音註</note></p>
  </div>
  <div org="uniform" sample="complete" part="N">
    <head> 唐紀一<note place="inline">起著雍攝提格(戊寅)正月，盡七月，
  <lb n="5771-100"/>不滿一年。</note></head>
  <p><dyn key="ch003" tar="c5771-025">唐</dyn>，古國名。<rmkey="r03948">
  <lb n="5771-125"/>於<m ref="c5771-025">唐</m>，其地<name>帝堯</name>、<nam
  <lb n="5771-150"/>太行</dm>、<rm key="2:r01347:d02746">恆山</rm>之西，<dm
  <lb n="5771-175"/>弼</name>等八人佐<dyn key="ch004">周</dyn>伐<dyn key="c
  <lb n="5771-200"/>受<name>魏禪</name>，<name>虎</name>已卒，乃追錄其功，封<
  <lb n="5771-225"/><m ref="c5771-140">&AS-8ADD;</m>生<name>淵</name>，襲圭
  <lb n="5771-250"/>因號曰<m ref="c5771-025">唐</m>。</p>
  <p> 高祖神堯大聖光孝皇帝上之上<note place="inline">諱淵，字<name>叔德
  <lb n="5771-275"/>本<rm key="3:r01758:r03975:d07714">隴西</rm><rm key="
  <lb n="5771-300"/>熙</name>家于<dm>武川</dm>；<name>熙</name>孫<name>虎</
  <p><milestone unit="date" ana="武德元年(戊寅、六一八)是年五月受隋禪，始改元
  <note place="inline">(<lb n="5771-325"/>戊寅、六一八)</note><
  <p n="1"> 春，正月，丁<lb n="5771-350"/>未朔，<dyn key="ch002">隋</c
  <lb n="5771-375"/>天子至于百官無不佩刀。<name>蔡謨</name>議云：大臣優禮者皆
  <lb n="5771-400"/>殿，非侍臣解之，蓋防刃也。近代以木，未詳所起；<name>東齊</

```

## インターネット時代の人文学の技術

— どのように処理し いかに表示するのか —

2004年9月6日(月)～9月10日(金)

場 所 京都大学人間・環境学研究科棟433・434演習室

定 員 約10名

セミナー内容

- 漢字情報学序説
- 多言語化概論
- XML, XMLプログラミング演習
- XSLT入門
- TEX, LATEX演習
- マークアップ概論
- 多言語情報処理概論

あなたは漢字が好きですか。それとももう漢字は要りませんか。漢字はいま大きなチャレンジを受けています。このプログラムでは、コンピュータと漢字のより良い関係を築くために、東アジア的規模でさまざまな方策を考えます。

(高田時雄)

2004年度

# 東アジア人文情報学サマーセミナー

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



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、ただか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。

## 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



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「Ω, 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 Ω 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。

## 実習2 LaTeX 演習



# ナー・カリキュラム

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

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.

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

In building on the basic knowlegde 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。

実習4 プログラミング演習