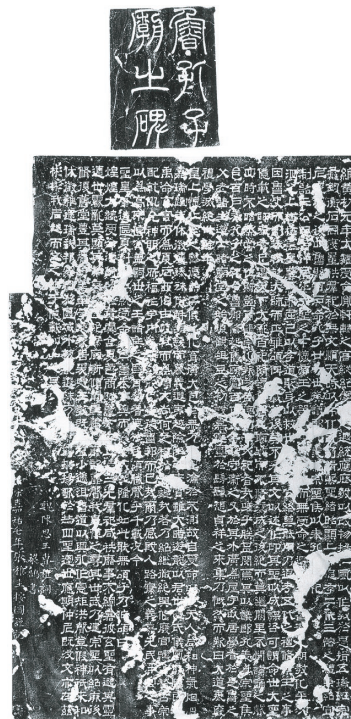


漢字と文化

漢字文化の全き継承と発展のために

京都大学 21 世紀 COE 東アジア世界の人文情報学研究教育據點

第 6 号



目次

アジャイルなソフトウェア開発	2
「第 2 回西夏学国際学術研究会」参加記	5
オントロジーな日々	8
COE Summer Seminar 2005: The World of XML Markup	11

大唐西域記序

橘寺

アジャイルなソフトウェア開発

——複雑さに抗して——

白須裕之

1. はじめに

読者の皆さんは、プロジェクトによってソフトウェアを開発するということに関して、どのようなイメージをお持ちでしょうか？ プログラミングをされたことのない多くの方が、おそらく工業製品の製造工程のようなものを想像されるのではないのでしょうか？ ソフトウェア開発はしばしば「複雑さへの挑戦」という文脈で扱われていますが、この複雑さはソフトウェア自身の複雑さ、要求・設計の複雑さであることもありますが、さらにソフトウェアを開発するチームの複雑さ、開発チームを囲む環境の複雑さといったことも考えられます。このようなソフトウェアにまつわる複雑さに、どのように向き合うかということについて考えてみたいと思います。

2. ハードウェアエンジニアリングとの類推

工業製品の場合、設計と製造の工程は厳格に区分されていて、設計の段階が終了して、製造チームに設計書が手渡されると、それをもとに大量の製品が製造されます。このようなハードウェアエンジニアリングの開発プロセスが科学たりえているのは、製造工程の品質管理にあります。それに対して設計の段階で使用されている手法は各製品ごとに異なったものになるではないのでしょうか。その場合でも設計チームと製造チームの境界にあるものは厳密に書かれた設計書です。

ソフトウェアエンジニアリングについても90年代ごろまでは、このようなハードウェアエンジニアリングを模範とした開発プロセスが主流を占めていました。設計チームが厳密な設計書を作り、プログラミングチームがその設計書をもとにプロ

グラムを開発する。品質管理は主に開発されたプログラムに対して行なわれ、多くの場合いろいろなテストが行なわれるといったものでした。このようなソフトウェアの開発手法がうまく機能するのは、全ての工程が予測可能である場合です。

このような考え方に根本的な反省をうながしたのが論文[6]です。ハードウェアエンジニアリングにおいて設計と製造の橋渡しをしている設計書は、ソフトウェアエンジニアリングの場合にはプログラムソースであるという主張です。すなわち、ソフトウェア開発には製造工程は存在しないのです。何故ならソフトウェアはプログラムソースから計算機がビルドしてくれるので、人間が手を下すことはないからです。このことはハードウェアエンジニアリングとの類推で行なってきた従来の品質管理、テスト、開発プロセスの考え方を根本的に変更してしまうことになります。

3. システムに対する要求

設計書がプログラムソースであり、ソフトウェアの開発には設計の工程のみあって、製造の工程がないとすると、ソフトウェアの開発プロセスをどのように考えれば良いのでしょうか？

まず、ソフトウェアの開発そのものについて考えてみましょう。ソフトウェアを作るには、解決すべき何らかの問題を抱えていて、その問題を解決するためにソフトウェアを必要とする人がいるはずですが、このような人を顧客と呼ぶことにしましょう。顧客がどのようにシステムを使って問題を解決したいかというソフトウェアに対する要求は、その顧客自身にも明確ではありませんから、これを明確化することから始めなければなりません。しかしソフトウェアを開発する前にこの要求

を明確化できるかという問題があります。開発中のソフトウェアを試しながら、顧客が要求を変更することもあります。すなわち、ソフトウェア開発は要求の明確化の過程であると捉えることができます。また、顧客のビジネス環境が時とともに変化し、問題自身が変化することも考えられます。このようにソフトウェア開発は予測不可能な変化を前提にして捉えないといけません。論文[6]が述べていることは、設計書、すなわちプログラムソースはビルド・テストによって何度も検証され、洗練していかなければならないことを主張しています。それではこの設計の工程がエンジニアリングたるためには、どのような開発手法が必要でしょうか？

4. アジャイルな開発手法

ソフトウェアの開発には本質的に顧客の絶えざる要求変更が伴います。ソフトウェア開発を成功させるために、人に関する問題を重視し、変化に対応できるための「価値と原則」がまとめられ、2001年に「アジャイルアライアンス宣言」[1]として発表されました。アジャイル agile とは「俊敏な」というような意味ですが、アジャイルな手法とは予測可能性を前提とせず、変化に俊敏に適応していくという適応性を重視した手法です。

この宣言に賛同する開発方法論には多くのものがあります。アジャイルアライアンス発足のきっかけにもなったXP (eXtreme Programming) が代表的な開発手法です。文献[3]がXPの原典です。XPは4つの価値と13の基本原則から構成されています。顧客を開発チームの一員とすること、短期リリースを行ないフィードバックを得ること、顧客による機能テストを行なうこと、設計を常に改善すること(リファクタリング)など、顧客の要求を明確化するための基本原則が含まれています。ひとつひとつの基本原則を取りあげてみると、実に単純なものであり、ソフトウェアを開発したことのある人であれば、良いと認めるようなプラクティスばかりです。しかし、それを組み合わせることで最大限に活用すると、適切な開発プロセスになるということが非常に特徴的です。もし

XPを実践されてみたいと思われる方はサイト[7]が参考になります。

5. 開発チームの進化

ソフトウェアを開発するチームを囲む環境は、予測不可能なカタチで日々変化していきます。このような変化に開発チームが適応していく過程は生物の進化に対比することができるのではないのでしょうか。文献[4]からの以下の引用(邦訳143頁)は、生物の形態デザインの改良について述べたものですが、「デザイン」を「開発チーム」と読みかえてみると、ソフトウェアの開発チームの歩みとして何か心に響くものがあります。

「デザイン」を絶え間なく改造するという過程は濃い霧の中を山に登るようなものである。頂上に着くという目的がなくても(あるいは、頂上がどこにあるかを知らなくても)、一歩ずつ上に行くという単純な決まりさえ守れば、一歩ごとに頂上に近づくことになる。

Darwinの「種の起源」出版から150年、進化の基本的な考え方は「自然適応」と「突然変異」だけで説明できるのでしょうか？この2つの概念だけで生物進化の全てを説明するためには、何か基本的な原理が欠けているようです。同様に開発チームがその環境に適応していくためには、チーム自身を進化させるための原動力、推進力なるものが不可欠ではないのでしょうか？

6. 適応型ソフトウェア開発

従来の手法では計画を守るために人間性が軽視されていました。アジャイルな手法は人間性を重視し、開発者と顧客の協調関係を重視します。XPではソースコードの共同所有、コーディング規約、最適ペース(週40時間労働)といった基本原則があります。しかし、どのようにすればチームを活性化させ、一人ひとりが生きいきと仕事ができ、創造性を発揮できるのかということについては、理論的に取り上げられていないような気がします。

このような問題に正面から取り組んだ開発手法に適応型ソフトウェア開発 (Adaptive Software Development) [5] があります。この開発手法もアジャイルなソフトウェア開発手法の一つですが、複雑系理論を拠り所とした点に特徴があり、以下の2つの概念を重視しています。

- ・ 適応は最適化よりも重要である。
- ・ 適者到達を特徴とする創発は、適者生存よりも重要である。

開発チームが唯一の勝者たらんとして最適化を試みることは、より高い山に登る機会を失なうこととなります (ここにもハードウェアエンジニアリングの最適化がふさわしくない例を見ることができます)。それよりも環境に適応していこうという到達の概念が重要です。「一歩ずつ上に行くという単純な決まりさえ守れば、一歩ごとに頂上に近づくことになる」のです。

しかし、そのように山を登り続ける原動力はどこから生れるのでしょうか？ それが創発です。創発 (emergence) とは複雑系理論の用語で、自律的な要素が多数集まることによって、その総和とは質的に異なる現象のことです。開発チームが進化していくための原動力は、このような個々のメンバーに還元できない創発的な秩序を如何に生み出すかにかかっています。適応型開発モデルでは思索、コラボレーション、学習のライフサイクルを主張していますが、ここでは学習についてだけ取り上げましょう (邦訳141頁)。

「トレーニング」とは、スキルや情報を習得することである。「学習」とは、心構えを作ることである。

学習をメンタルモデルから捉えるのがこの開発モデルの特徴です。

7. 終りに

ソフトウェア開発はただプログラムを作れば良

いというものではありません。むしろプログラムの作成はソフトウェア開発のほんの一部でしかないのです。ソフトウェア開発とはその全過程が顧客の要求を明確化していくプロセスであると言えます。アジャイルな開発手法は「身軽な旅」を標榜して、困難な道を歩みきろうという、そんなプロセスです。プロジェクトに関わるメンバーの思索、コラボレーション、学習の結果が生み出す人間的な活動です。そこにはチームの進化が欠かせません。現在、我々は唐代人物知識ベース、唐代官職知識ベースの設計・開発に従事しています。これら知識ベースの研究開発を通してチームがどのように進化していくのか楽しみにしています。

参考文献についての補足

オブジェクト指向の考え方を使得アジャイルなソフトウェア開発を実践したい方には、文献 [2] が参考になります。

原稿を読んでいただいた秋山陽一郎さんに感謝いたします。助言を十分活かせなかったのは筆者の責任です。また、我ら開発チームの進化の原動力である「まつお館」の皆様にも感謝いたします。

参考文献

- [1] Agile Alliance, <http://www.agilealliance.org/>
- [2] S. W. Ambler, Object Primer, 3rd edition: Agile Model-Driven Development With UML 2.0, Cambridge University Press, 2004. (邦訳: スコット・W・アンブラー, オブジェクト開発の神髄, 日経BP社, 2005)
- [3] K. Beck, eXtreme Programming eXplained, (邦訳: ケント・ベック, エクストリーム・プログラミング入門, ピアソン・エデュケーション, 2000)
- [4] B. Charlesworth and D. Charlesworth, Evolution: A Very Short Introduction, Oxford University Press, 2003. (邦訳: ブライアン・チャールズワース, デボラ・チャールズワース, 進化, 岩波書店, 2005)
- [5] J. A. Highsmith III, Adaptive Software Development, 2000, (邦訳: ジム・ハイスミス, 適応型ソフトウェア開発, 翔泳社, 2003)
- [6] J. Reeves, What Is Software Design?, C++ Journal 2(2), 1992. <http://www.bleeding-edge.com/Publications/C++ Journal/Cpjour2.htm>
- [7] D. Wells, Extreme Programming: A Gentle Introduction, <http://www.extremeprogramming.org/>

「第2回西夏学国際学術研討会」参加記

池田 巧

2005年8月16日より4日間の日程で第2回西夏学国際学術研討会が寧夏回族自治区の銀川市にて開催された。1995年夏の第1回から、実に10年ぶりの開催である。西夏はよく知られているように元の統一以前の西北中国に党項（タングート）と呼ばれる民族が建てた国で、漢字をモデルに独自の西夏文字を制作し使用していた。発掘された文書は20世紀に中国・日本・ロシア・台湾で精力的に解読が進められ、6千余字のかんりの部分の意味が判明している。現在では西夏文献研究を通じて当時の漢文文献には記録されなかった社会のさまざまな側面が解明されつつある。

国際学術研討会という名称ながら、外国からの参加者はわずかにロシアから1名、日本から3名であり、ロシアのソローニン氏（歴史学）と日本の小高裕次氏（言語学）は現在、台湾で仕事をしているので、完全な国外からの参加者は、当初私と北室南苑女史（書家）の二人のみであったが、最終日に東京外大 AA 研の荒川慎太郎氏（言語学）、大阪大学研究員の佐藤貴保氏（歴史学）、地球環境学研究所研究員の森谷一樹氏（歴史学）の3名が駆けつけた。いっぽう中国国内からの参加者は250名にも及ぶ大規模なもので、白濱氏（社会科学院民族研究所）、牛達生氏（寧夏文物考古研究所）、台湾からは龔煌城教授と林英津女史といった著名な研究者のほか、地元の企業人や政府関係者をはじめ、党項羌は同族だとばかりに四川省のアパ自治州から来た羌族の代表団、我こそは西夏の後裔だと名乗る洛陽の郷土史家、西夏文字で創作活動をしている書家や芸術家などの多彩な顔ぶれが参加していた。銀川市では現在最高級の4星ホテルである太陽神賓館を会場に、最近の中国で開催される学会の常として、協賛にいくつも



観光地として整備された現在の西夏王陵（三号陵）

の企業がつき、観光産業の活性化への貢献が期待されているのが一目瞭然であった。平たく言えば「西夏で寧夏の町興し」である。最終日には観光会社の賛助による西夏王陵、賀蘭山の岩絵、沙湖の砂丘への観光が組まれていた。

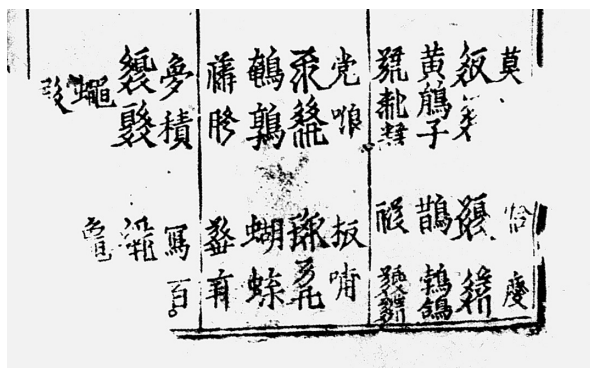
研究発表は、歴史と文化、言語と文献、考古と他の3つの分科会に分かれて、各分野とも2日間でそれぞれ20名を数える多数の報告があった。初日の分科会で行なった私の研究報告には、最近あちこちの研究会でよく一緒になるナシ族の友人の木仕華氏（社会科学院民族研究所副研究員）から、いつもながら細かい点にこだわったコメントと関連事象にかんする長い補足説明を加えてもらった。おかげで他からの質問を受ける時間がなくなり、内心しめたと感謝さえしたものである。安心して締めくくろうとしたら、司会の聶鴻音氏（社会科学院民族研究所研究員）から、ちょっといいか、と短いけれども本質をつく有益なコメントをいただいた。西夏語との関連が注目される中国四川省の少数民族言語のムニャ（木雅）語で‘蝶’は / mbē mbu / と発音され、西夏語と同源語で

ある可能性が高いと報告した事例について、最も信頼できる龔煌城教授による西夏文字の再構音では *pja pju となってしまうけれども、西夏の口語語彙を記録した対照語彙集の『番漢合時掌中珠』(1190年刊)の音訳漢字は「板哺」であり、こちらのほうがムニャ語の発音により近く、その親近性を彷彿とさせ、西夏語音に対する忠実な音写がなされていた可能性があるのではないか、という指摘であった。さすがに鋭いと思い緊張したけれども、幸いに想定範囲内の問題だったので、さらりと見解を述べて謝辞で締めくくり、数ヶ月におよぶ研究は、わずか10分弱の報告と7分の質疑応答(そのうち5分は木仕華のコメント)でほぼ制限時間ちょうどに終了した。この機会にぜひご意見を伺いたかった龔煌城教授は、長旅の疲れで気分がすぐれず休んでおられた由で、分科会の会場におられなかったのが残念でならない。

研究発表を除くと、学会の話題の中心は、寧夏における西夏研究の立場の主張と国内外の西夏文献資料の出版動向であった。主催者代表の李範文教授による演説とも言うべき報告は、西夏の故地である寧夏こそが西夏研究の中心として当然その責務を担うべきであり、しかし文献や文物の多くは持ち去られて本拠地には何も無いというジレンマ、それゆえ『俄藏黑水城文献』につづく『英藏西域文献』『中国藏西夏文献』の大型シリーズの刊行は、誰もが国内外の西夏文献を利用し得る多大な便宜をもたらすものであり、大いに歓迎したいという主旨であった。出版社の責任者の紹介と

簡単なスピーチもあったものの、刊行に到るまでには幾多の問題が生じていたらしく、李範文教授はキーノートスピーチに匹敵する長時間を割いてさまざまな事情を語り、顔を真っ赤にして吠えるかのように西夏研究の現状と無理解への不満を爆発させていたのには驚かされた。

私と小高氏はこうした資料の出版を機に、西夏文字のデジタル化についての議論があるものと期待していたのだが、残念ながら学会のテーマとして取り上げられてはいなかった。文献資料の出版の次の段階として、デジタル化による研究成果のデータの共有や相互利用の発想がないとすれば、いくら寧夏が研究の中心を担うと主張しても、実質的な求心力を得るのは難しいのではないか。小高氏とそんな意見を交わしていたところ、情報化に関して「西夏文録入系統」という新しいソフトウェアの紹介があり、研究成果の編集出版にもすでに使われていて、フォントは美しく新発見の文字や異体字まで収録したので、今後はこれを大いに使って行きたいという旨の報告があった。商業ベースに乗らない西夏文字フォントの制作は、たいへんな事業であり、仕様を如何に定めるかがきわめて重要であるだけに、私は期待と羨望と不安がないまぜになった心持ちでそのアナウンスを聞いていた。けれどもあとで開発者から恵贈を受けたこのフォントと入力システムが、中国語簡体字版 Windows95 / 98 / ME 専用(2000 / XP は不可)、しかも「方正」コード上に外字として展開し、入力には6桁の四角號碼番号をローマ字キーに置きかえて行なう、と知って啞然とした。世界各国の研究者が自ら利用できる仕様にはほど遠いもので、「ぜひ使ってみてご意見を」と言われたけれども、これでは使いようがない。せっかくの西夏文字の情報化も出版の便宜に留まり「ワープロはオペレータまかせの清書機械」の発想から脱却していない。その旨を開発者と話そうか、とも思ったが、祝賀ムードの学会のロビーで立ち話をしたところで、せっかく始めた事業に水をさしているだけと受け取られかねないと思ってやめにした。数年前、台湾で開発した西夏文字フォントと入力法について話を伺う目的で中央研究院を訪問した



『番漢合時掌中珠』第17頁に見える「蝴蝶」の対音
(『俄藏黑水城文献』第10巻所収)

西夏文録入碼検索表

210144—212152

1951	𐌪𐌰	eyoeae	1974	𐌪𐌰	eyoxxf	1997	𐌪𐌰	eyeyab
210240	𐌪𐌰	eyoexo	210444	𐌪𐌰	eyoxxx	1998	𐌪𐌰	eyeyac
1952	𐌪𐌰	eyoexo	1975	𐌪𐌰	eyoxxg	1999	𐌪𐌰	eyeyad
210242	𐌪𐌰	eyoexe	1976	𐌪𐌰	eyoxxh	2000	𐌪𐌰	eyeyae
1953	𐌪𐌰	eyoexa	210450	𐌪𐌰	eyoxwo	2001	𐌪𐌰	eyeyaf
1954	𐌪𐌰	eyoexb	1977	𐌪𐌰	eyoxwa	2002	𐌪𐌰	eyeyag
210244	𐌪𐌰	eyoexx	1978	𐌪𐌰	eyoxwb	2003	𐌪𐌰	eyeyah
1955	𐌪𐌰	eyoexc	1979	𐌪𐌰	eyoxwc	212125	𐌪𐌰	eyeyew
1956	𐌪𐌰	eyoexd	210452	𐌪𐌰	eyoxwe	2004	𐌪𐌰	eyeyew
210250	𐌪𐌰	eyoewo	1980	𐌪𐌰	eyoxwe	212140	𐌪𐌰	eyeyxo
1957	𐌪𐌰	eyoewo	210545	𐌪𐌰	eyoxwv	2005	𐌪𐌰	eyeyba

『西夏文録入系統使用手冊』の検字表の一部

とき、開発者から Big5 上の外字処理で何でもできるのだからとにかくこれを使えばいい、と一方的にまくしたてられ、当時抱いていたデータの共有についての私の見解やアイデアなどは、全く聞かずにもらえなかった苦い記憶が頭をよぎったからである。

開発者からは「ぜひ専門家の意見を伺いたい」とも言われたが、彼のいう「専門家」は目の前の私のことではない。会場では知己も挨拶を交わした初対面の研究者も私が日本の京都大学からの参加だとわかると、例外なく私のことなどさておいて、すぐさま高名な教授の名前を挙げ、その消息を訊ねるのが常であった。ここで西夏文献研究を専門としない私が、理系の技術者に向かって「西夏文字のデータ処理についてはオレがその専門家だ」と宣言して議論を始めたとしても、所詮は「誰だか知らない外国のお客さんからの個人的な参考意見」として相手にされなかったことだろう。

西夏文字の情報化については、上述の「西夏文録入系統」のほか、これまで今昔文字鏡 [http://www.mojikyo.org]／東京外国語大学アジア・アフリカ言語文化研究所 [http://www3.aa.tufs.ac.jp]／台湾中央研究院 [西夏文字情報は内部利用で一般には非公開]／中華西夏文処理系統 [市販品：北京中易中標電子信息技术有限公司 http://www.china-e.com.cn] があり、小高裕次氏が個人で制作しているフォント [http://teacher.wtuc.edu.tw/92029/html/index.html] もあるが、すべて独自仕様で開発されたもので、

互換性は全くない。いずれも外字もしくは漢字の被せフォントである。とはいえ被せフォントなら、正体は漢字のテキストファイルなのだから、下になる漢字と被せる西夏文字との間での対応を統一しておけば、JIS と GB と Big5 で漢字をコンバートしても、被せる西夏文字は同じにできるのでデータの共有が可能になる。そして日本語ベースでも中国語ベースでも、IME の辞書登録をうまく利用すれば、ユーザーがそれぞれ自分にあった西夏文字の入力システムを工夫できる。このアイディアを実現すべく安岡孝一さんにご協力いただいて、ユニコードで JIS と GB と Big5 で同じ文字と認定され、包摂した CJK 漢字のデータまで準備してもらったのだけれど、文字鏡にも中央研究院にも理解を得ることはできなかった。それぞれに相当な労力をかけて制作された字形データを建設的に未来に活かすためには、いまこそアジアの文字の万国公法を説く坂本龍馬の出現を望みたいと思うのは、私ひとりだけであろうか。

学会終了の翌日、銀川空港でチェックインを済ませると、北京行きの飛行機の出発が少し遅れるという場内アナウンスがロビーに流れた。同じ便に乗る待合室で龔煌城先生は、池田さん、前からずっと気になっていることがあるんですが、と温和な表情でおっしゃった。♪祝う今日こそ楽しけれ…。で「楽しけれ」となるのは「こそ」との呼応による活用形だから、という問題に始まって、係り結びが唱歌や童謡の歌詞の中に今も活きていることや現代の口語と古典語との対応の例などをあれこれ語りあったのち、先生はふと、♪里わの火影も森の色も、と口ずさんでから、ところでこの里ワのワってどういう意味でしょう？ 辞典もいろいろ引いてみましたが、よくわからないのです、と真剣な表情になられた。さて、それは…と答えに窮したところで再び場内アナウンスが流れ、慌ただしく搭乗の準備が始まった。機内では席が離れて楽しい議論も中断したまま北京に到着。北京空港で厦門行に乘換える龔煌城先生ご夫妻と別れの挨拶をしたあと、私には日本語の解明が宿題として残された。

オントロジーな日々

永田知之

筆者が京都大学人文科学研究所に附属漢字情報研究センターの助手として採用されたのは、今年、即ち2005年5月のことであった。「唐代文学の専攻者でデータベース構築に関心を有する者」という、前半と後半にあまり関連性の無い、奇妙といえば奇妙な公募条件を2月の初めにネット上で閲覧したのが、今となっては遠い昔のことにように思われる。

学部・大学院で通計九年間（二年の中国留学期間を含む）も京都大学の中国文学研究室に籍を置いてきただけに「唐代文学の専攻者」と自称するほどの厚かましさは既に持ち合わせていた。問題は「データベース構築に関心を有する者」の方である。データをあれこれ集めるのは好きな性分だが、それによって何かを作りあげる意志や能力をもっているとはお世辞にもいえそうにない。そもそもコンピュータに関する知識を致命的なまでに欠いていることは、本人が誰よりもよく自覚していた。「次の助手はいま人文研でやっている電算機関係の事業に従事するらしい」という噂を聞いていただけに、応募するのはややためらわれた。

幸いにというべきか、面接試験ではそれほど突っ込んだことも聞かれず、研究所に採ってもらえた。そして着任と同時に「21世紀 COE プログラム」の一環と位置づけられている「漢字文献ナレッジベースの構築」の責任者（？）らしき役割を担うことになったのだ。

入所した当座の三週間ほどはできるだけ毎日、研究所分館から自転車で五分内外の「まつお館」へ足を運ぶようにしていた。そこで半日を過ごすこともそう珍しくはなかった。

「まつお館」は今出川通に面した中華料理店にして、チャンポンで名高い「まつお」方の三階に

あるテナントを借りて設けられた COE の基地である。フローリングで五十畳ほどある部屋にパソコンやその他の事務機器、生活用品、必要資料が運び込まれ、研究員（3名）、研究補助員、アルバイトの諸氏で多い時には、10名以上の方々が勤務している。その業務内容はナレッジベースのみならず、「CHISE プロジェクト」等 COE 全般に涉っている。

商店や民家が建ち並ぶ街中の一角で、「東アジア人文情報学」という新たなる学問領域が生み出されつつあるとは「お釈迦様でもご存知あるめえ」と思われる。それにしても「チャンポン」という日中の食文化を融合した料理で知られる店の真上で非アルファベット文献と電算技術の融合が図られている、この状況を示唆的に思うのは筆者だけであろうか。

大した用も無いのに平日はほとんど休み無くやって来る新参者を、「まつお館」の皆さんは寛大にも受け入れて下さった。基地が形成されてから既に一年半、完成した人間関係の中に入っていけるか、柄にもなく心配していた筆者にとって、これは望外の幸運であった。

さて、本プロジェクトの目的だが、その詳細については、本誌創刊号所載のウィッテルン・井波両先生による「唐代研究ナレッジベース」を参照されたい。つまるところ、データをコンピュータに入力し、人間には真似のできない情報処理能力で、それを中国学研究へ生かそうということである。「唐代の著名人で A 地方の出身者にはどういう人物がいるか」、「B という一族について、いま分かる限り詳しい家系図を見たい」、「C という官庁にある時期、どのような人々が所属してい

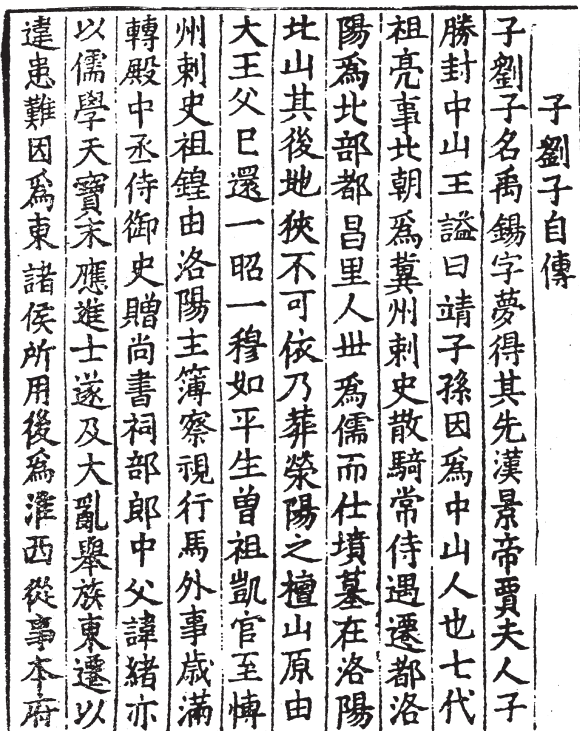


図1 唐・劉禹錫「子劉氏自伝」

たか」，そういった問いや希望の答えが，簡単なキー操作で，瞬時にディスプレイ上へ表示される「夢の機械」を造ろうというわけである。

しかし，この「夢の機械」への道のりが全く平坦ではないのだ。文献上の情報を計算機が理解且つ処理できるよう型（フォーマット）を決める，モデリングという作業がそこでは不可欠となる。時に研究員相互間で人名に関して「父の死後，母の再婚により姓だけでなく名も改めた人物が，後に任官試験に合格した折，皇帝（それも地方政権の）のお声掛けで元の姓に戻り，名前を新しく改めたという内容をどの情報も落とさずにコンピュータに分からせるにはどうすればいいか？」という一見とんでもない議論が始まったりする。

「こういうケースは例外でして……」，おずおずと口を挟む筆者に「そういう例外的なケースを最初から想定しておかなくてはならないんですよ」と研究員某氏が静かに，だが断固とした口調で教えを垂れて下さった。いかにも，紙の上の論文・レジュメならば書き終わってからでも，何か思いついたら，特異な例として注釈にでも入れておけ

- <関係地>
- <場所 type="貫籍">
- <区域>洛陽</区域>
- </場所>
- <場所 type="出自">
- <区域>中山</区域>
- </場所>
- <場所 type="出自">
- <区域>彭城</区域>
- </場所>
- <場所 type="生地">
- <区域>江南</区域>
- </場所>
- <場所 type="卒地">
- <区域>洛陽</区域>
- </場所>
- </関係地>
- </個人>
- </人物>

図2 図1に見える情報の一部をXML文書で表現したもの

ばよいかもしれない。だが，計算機の場合はそうはいかない。最初からその種のデータを組み込む余地を準備しておく必要がある。例外的なものを後で追加するのは容易ではないのだ，云々と。

6月以降は「まつお館」のスタッフたちと週1～3回のペースで1回2，3時間のミーティングを重ねてきた。筆者と同世代で20～40代の情報工学など異分野を専攻する人々との議論には，常に蒙を啓かれる。全くの未経験者相手に噛んで含めるようにものを教える彼らの苦労は並大抵ではあるまいが。

「ある人物の伝記で『彼は知制誥，郎中，中書舎人，侍郎を歴任した』と書いてあったら，それはこれらの官職を順番に経験してきたと考えない方がよいでしょう。知制誥は他の官職にある者が中書舎人を助ける時に帯びる官名だから，恐らく郎中兼知制誥というのが実情。そのあと中書舎人，侍郎へ昇進していったという情報がここから読み取れるんです」。制度史の専門家が聞いたら呆れかえりそうなほどいい加減な説明を，筆者がした際のことであった。「多数のデータを集めれば人



図3 「まつお館」の内部

間に代わって計算機が高い確度で兼任や昇任の過程を推論してくれます。これがコンピュータ・サイエンスにおけるオントロジーというものです。是非やりましょう」。新たな課題の出現に対して研究員某氏の目は輝いていた。

「そいつは便利だなあ」と思う反面、文献学に携わる者の端くれとしては「ちょっと待ってくれよ」という思いも頭をよぎる。表面上の個別な記録からは読み取れない官制における法則性といった事柄は、諸先学が血の滲むような苦労の末に発見してきたものだ。「眼光紙背に徹する」職人芸めいた文献読解能力があってこそ、それは可能だったといえる。

その成果をより確実な形で、誰もが一瞬にして手に入れられるようになる。今日明日に我々がそういった段階へ到達するという話ではない。だが、それが実現した暁には中国学の変容は避けられないだろう。得られるものは大きいが失うもの（例えば「職人芸」を培う機会など）も存在するのではないか。

時代が変われば学問も変わる、その程度は頭で理解しているつもりだった。そこに戸惑いを感じるのは筆者がアナログ派であるためか、或いは文献学が文献学たる所以に何かしら変化が起きているためか、それは今後、作業が進行する中で明らかになるかもしれない。もし、明らかになるとしたら、それこそナリッジベース以上に大きな成果ともなり得よう。

入所以後、「永田君、しんどくないか」と所内

の何人かの方から筆者の体を気遣う言葉を掛けていただいていた。「全くしんどくない」といえばそれは嘘だが、何が見えてくるか（或いは見えてこないか）分からぬ作業を、生意気な言い様ながら、当人は少し楽しみ始めている。

本業のコンピュータ関連だけで仕事が山ほどあるにも関わらず、自ら進んで礪波護著『唐の行政機構と官僚』（中公文庫）を読み、ナリッジベース作成に役立つ知識を得ようとする研究員諸氏の勉強熱心さには頭が下がるばかりである。その姿に申し訳無さを感じたのが直接の動機となり、筆者の方も（株）テクノロジックアート著『独習UML』第3版（翔泳社）に目を通すことにした。三百頁を超える同書にここ数ヶ月の間、悪戦苦闘しているが、今もって Unified Modeling Language（統一モデリング言語）とは何のことかよく分からない。

かくて、夏から秋にかけ筆者はまごまごしているばかりだったが、「まつお館」スタッフ各位の献身的な努力により、このプロジェクト最初の成果となるべき人物データベースの入力は既に終了した。次の課題とされる官職データベースのモデリングと並行して、11月中旬現在その実装作業を進めつつある。人手が足りない分、こういった二正面作戦もやむを得ない。

いずれにせよ、今はせいぜい分館と「まつお館」の連絡係でしかないが、筆者としては周囲の人々において行かれぬよう、微力ながら自らを向上させていかなければならない。「オントロジー」という言葉が「おっとろしい」か「おっそろしい」にしか聞こえなかったと素人（表題のオントロジーは「対象とする世界に存在するものごとを体系的に分類し、その関係を記述するもの」という意味で用いている、念のため）を責任者に採用したことが人文研の人事史上における最大の失敗とならぬためにも、研究所内外各方面から我々の業務へ、今後なお一層のご支持、ご支援を賜るようお願い申し上げて、筆を擱くこととしたい。

COE Summer Seminar 2005 : The World of XML Markup

Christian Wittern

Introduction

The digital medium, as it has established itself in the last decades, forces a completely new, radically different view on its users, even if the content is the seemingly familiar one of texts, well known as they are from their long existence on surfaces that carried their forms, such as bones, bamboo, silk, stone, wood or paper. In the digital medium, there is no surface any more, the text has to be decoded and newly digitally encoded. While the implications of this are just starting to being explored, there can be no doubt that they will change substantially the way we interact with text. It seems therefore a necessity of some urgency to allow young scholars who have a principal interest in texts and its tradition, to explore some of these implications and reflect on how this might reshape their scholarly agenda. It is with this mindset that the introduction to text processing based on XML, which is one of the most advanced and most frequently employed basic technologies for dealing with structured data, has been conceived.

The time available for this introduction did of course only allow some very preliminary exposure, so the design of the curriculum was of especial importance. It had to contain some introduction to the historical background against these technologies developed, but also cover the main features of XML as a generic

data encoding technology. On the other hand there should also be a practice session and in order to give students enough insight as to enable them to continue on their own, an introduction (or maybe more a nod) to the *Guidelines* of the **Text Encoding Initiative** was also given. To counter the danger of intimidating the students right from the beginning, there were also some real world examples and testimony of applying this technology to Yuan Dynasty documents within the context of a research seminar, courtesy of Professor **Iwai Shigeki**. All in all, judging from the feedback and general atmosphere in the classroom during this intensive week, it seems to have been quite a good mixture.

What is text, really ?

The program started out with a reflection on the underlying assumptions and theories concerning text, then proceeded to consider the implications this has on the technologies





used for processing digital text.

Depending on the context of its usage, text can be:

- In everyday usage, a broad term for something written to express something.
- In linguistics, a communicative act, fulfilling the principles of textuality.
- In literary theory, text is the object studied, be it a novel, poem, film, advertisement or anything else with a linguistic component. This broad use is inspired by semiotics and cultural studies of the 1980s.
- In information processing, text refers to character data.

Text can be very simple or complexly structured. Structure usually makes it easier to understand. Text depends on some notation, usually a script made up of characters.

Text encoding is the process of transcribing a text in digital form. It is sometimes confused with character encoding, which is the act of assigning characters to the individual items observed in the stream of text. Text encoding comprises character encoding, but goes beyond that, since it is also concerned with re-creating the structure of a text in electronic form. Text encoding is sometimes also confused with markup, which is a methodology used in text encoding to express information about structure, status or other special features of a text.

The term markup derives from the annotations or symbols editors would use in traditional publishing to convey to the printers information about how the text should be printed: Size and weight of characters, position on the page, size of margins etc.

Some theorists of markup have generalized the notion of markup to include certain aspects of written communications, for example page layout, typography and punctuation.

Markup of digital text is usually classified in different categories according to the way it is used: presentational, procedural, and descriptive.

XML as an enabling technology for text processing

In the following sections, the participants were made familiar with the technical details of the XML technology. Mastering the underlying principles and its consequences is essential for successful application. They are also not arbitrarily and not simply technical details that can be ignored, so some time was spent here in familiarizing the participants with the XML model of a text, which is basically a tree structure and what implications this has for the ways it can be used for text processing.

At the same time, details of working with XML and an introduction to the syntactical



forms were introduced while working with a specialized XML editor, which allowed the students to immediately appreciate different views on a text, as well as simple XML technologies such as addressing certain parts of a text with XPATH, which are far superior to locate and collect places of interest in a text than traditional searching based on string matching. This concept was well received with the students and contributed to overcoming their initial reservation at working with the 'source code' of XML.

Recording results of analytic reading processes with XML

An important part of this seminar were the practice sessions, which were designed to provide participants a first hand experience of how markup can be used to record the reading process, which consists of making analytic discoveries about the constituent features of a text. As a sample, we prepared a section from *Sima Guang's* 司馬光 *Zizhi tongjian* 資治通鑑。Since the most commonly used editions today incorporate his separately compiled 'critic of sources' *Kaoyi* 考異 and also the commentaries by *Hu Sanxing* 胡三省 and others, the structure of this text is rather complex. For this reason, we reserved this version for the second practice session and started the first session with an earlier woodblock print, which contained just the text, without commentary.



We asked the participants to mark the structural units of the text, discernable through the spaces, but also names of persons and places. Since we could not expect all participants to be able to perform this analysis on their own, they were also handed a color-coded 'cheat-sheet', that indicated the proper names in the text using different colors. Although the time reserved for this session was rather short, a few participants did indeed, much to our surprise, finish the whole text of the section.

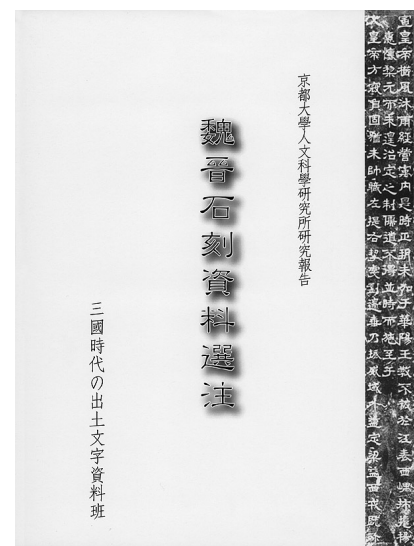
Epilogue

It goes without saying that the few hours set aside in this seminar for an introduction to XML based text encoding can not by itself provide enough time and exposure to enable a full familiarity necessary to allow for independent work, but its goal has been achieved if it provided the participants with enough taste of it to wet their appetite for more.

出版

- ・三国時代の出土文字資料班『魏晉石刻資料選注』京都大学人文科学研究所，2005年が出版されました。

「ただ、この事業はたんにインターネット上で所蔵資料を公開するために企画されたものではないことを、とくに強調しておかねばならない。舊東洋学文献センターを拡充改組して二〇〇〇年四月に発足した漢字情報研究センターの基本戦略，すなわち漢字文献の公開およびその国際的情報交換を円滑に行うために適正な漢字管理システムを開発するという目的に沿って，中長期的な見通しの下で行われているのである。たとえば，実際に彫られた文字の釋讀および語注は，適切な検索システムが構築されたあかつきには，デジタル画像とリンクされ，利用者の便宜を飛躍的に増大させることになるであろう。そのためには，あらかじめ共同研究による精密な譯注作業を行い，十分な検討を重ねておく必要がある。」（井波陵一教授による前文より抜粋）



本報告書は近日中に PDF ファイルとしてウェブ上に公開される予定です。

TOPICS

シンポジウム，セミナー

- ・ 2005年9月5日～9日 京都大学人間・環境学研究科棟において 2005年度東アジア人文情報学サマーセミナー「インターネット時代の人文学の技術（スキル）―XML 世界への誘い―」を開催しました。
- ・ 2005年10月13日～14日 京都国際交流会館において「CHISE Conference 2005」を開催しました。
- ・ 2005年10月14日～15日 京都大学人文科学研究所において「CodeFest 京都 2005」を開催しました。
- ・ 2005年11月18日 京都大学人文科学研究所において 国際ワークショップ「近代東アジアの情報―質と量」を開催しました。

以上のセミナー，シンポジウム等の成果は本プログラム・ウェブページ (<http://coe21.zinbun.kyoto-u.ac.jp/>) 上に公開される予定です。

Chinese Characters
and Culture



発行日 2005年11月30日
発行者 文部科学省21世紀 COE プログラム
「東アジアにおける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/>

