Knowledgebase of Tang Civilization project in Kyoto[1]

Christian Wittern
Kyoto University, Institute for Research in Humanities

Table of Contents

Introduction

The projects described here have started as part of the COE program “Toward an Overall Inheritance and Development of Kanji Culture” with funding from 2003 to 2008.

There are two main objectives: The first one is to develop methods utilizing information technology to make not just the words of the documents that serve as witness to traditional Chinese culture, but also the underlying structures of knowledge that are implicitly embedded therein, accessible for analytic processing. The second objective is to build a prototype of such a database,

[1]This is a preliminary report not for citation.

in this case, documenting China, its culture and intellectual mindset as of the Tang period (618-907).

In the presentation, I will focus on two projects, that have been conducted as part of this:

- Resources for Tang Studies
- Knowledgebase of Tang persons

Both of these do collect material about Tang persons, but the way this is done is quite different. The overall objective however, that is, the attempt to represent such information in a machine readable form, that can be used to further analyse and process the information is similar. I will start out with an excerpt of the plan, as of 2003, which in hindsight is of course to be taken with a grain of salt.

Resources for Tang Studies

The following is an excerpt from a document that laid out the plan for the Knowledgebase part of the project[2]:

> The Knowledgebase of Tang Civilization (唐代研究ナリジベース) is an attempt [...][to provide] a comprehensive electronic archive of information about China during the Tang dynasty in a way that allows new ways to access, analyze and expand the information. It's main point of access for researchers will be a web application, but other interfaces will be developed.
>
> Initially most of the information will be textual, enhanced by images, visual reproductions of objects, such as digitized maps. The distinguishing feature of the knowledgebase is the way information items are interconnected in a flexible and innovative way.
>
> The information in the knowledgebase will be organized along the following information axis:
>
> - Personal names, dates and activities of people of the Tang.
> - Placenames and references to locations, administrative geographical units.
> - Works created during the Tang, including texts, artefacts and buildings
> - Calendar and time
> - Events of importance and influence

[2]The quotations are from an English draft that has never been published; a Japanese version translated by Yōichiro Akiyama 秋山陽一郎 has been published as Wittern 2004.

Work started out with the two dynastic histories and the section on Tang of the *Zizhi tongjian* 資治通鑑 by 司馬光 Sima Guang (referred to as *Tangji* 唐記 from now on). Early on, it was decided to work from the modern punctuated editions, since problems of textual transmissions seemed to play a minor role for these texts, especially when weighted against the gains that could be reaped by taking advantage of the richness of information that could be inferred from the punctuated text, and was such available for processing, of which in the context of this project, the markers for proper nouns where especially interesting.

During the input process, which was outsourced, we thus asked the keyboarding company to add a number of extra symbols for the proper nouns, which could at the input level be divided into two classes, titles of works on the one hand, and other proper nouns (including names of persons and places, era names etc.) on the other hand. It was then attempted to disambiguate the proper names and assign them to the corresponding categories. At the same time, the adding of structural markup according to the Guidelines of the Text Encoding Initiative (TEI) to the texts began.

To proceed with the tools available turned out to be rather difficult. First attempts at using simple heuristics for automatic disambiguation of terms where not very satisfying, so it seemed to be more promising to manually work and then use the result as a training base for automatic procedures. Of the different texts under consideration, the *Zizhi tongjian Tangji* was selected for manual markup of proper nouns. In a first phase, special attention was paid to personal names. For this purpose a registry of persons mentioned in the texts was started. Every person was given an entry in this registry and assigned a key code that could be used to identify this person. This key was then entered into the texts whenever this person was mentioned, whatever the name used. This required therefore a considerable effort to identify persons, especially where different persons with the same name exists, or where a person has been addressed in the texts with a variety of names. In addition, since the work on this was divided up among different operators, lists of names and persons had to be maintained and made available for reference, which again was difficult with the tools at hand. However, the advantage of this procedure clearly was that only attested names were entered in the databases, and all entries where directly linked to their sources in the texts, so that verification was possible at every stage. Only with the advent of a completely Unicode compliant version of the (commercial) database development program Filemaker became it feasible to use this for maintaining some of the tables.

To make this process of identification easier, in addition to the texts already available for reference, we compiled lists of persons and their names from available sources that could be used as an online reference. At the moment, this reference file has grown to almost 30000 persons. This

reference file will be reusable for other purposes as well, for example in assisting the analysis of texts that have not yet been marked.

While the main work was concerned with persons, some work has also done to assign identifiers to place names and track the changes that took place here over time. Since, contrary to persons, the size and location of administrative units also changed, this is a much more difficult task that requires more work to become useful. A lot of work in this area, based on much better expertise was already under way elsewhere, so it seemed better to avoid duplication of effort and rather look for opportunities to eventually integrate our work with databases in existence.

Conceptually, the content of the repository can, as outlined above, be divided into the primary resources, which are the source texts in this case, on the one hand, and additional information about the text, its content, persons and places mentioned in the text, events that occurred and all the things mentioned above. There is considerable debate about how such information could/should be handled. One school of thought seeks to organize this additional information into hierarchies of classes and their instance, in a way similar to philosophical classification systems as proposed for example by Aristotle. This will allow to specify very general and abstract questions like "show me all sections that mention birds" or "show me all persons born in the coastal areas"[3]. The other school of thought tends to organize things in a less structured, ad-hoc way by attaching labels to information items, but not using these labels for purposes of organizing these items in a hierarchical way. This way of organizing information has recently been called folksonomy (as a combination of folklore and taxonomy) since it is frequently connected with discussions of the "Web 2.0" [4].

In addition to these approaches, a more theory neutral approach to encoding such information items based on Topic Maps[5] has also been tried[6] but in the end it turned to be too inflexible and ineffective for the purpose.

Results

As originally planed, the texts that have been proofread and marked will be available through a web application interface developed for this purpose, more details of this will be given below. In

[3]The efforts surrounding the Semantic Web, do belong here, see for example Berners-Lee, Hendler and Lassila (2001) and Berners-Lee, Shadbolt and Hall (2006), but also the discussion of ontologies and other efforts inheriting methods from AI (Artificial Intelligence) processing in a more general sense.
[4]Web 2.0 is a recent buzzword, which was introduced to the larger public by the publisher Tim O'Reilly in to indicate a new modus of the World Wide Web that allowed more user participation; among the characteristics mentioned by O'Reilly is the use of tags (folksonomies).
[5]A method to associate arbitrary information items with information ressources, and in addition ot that also encode relations among these information items. A international standard (exists for Topic Maps and a small community of vendors and developers is developing applications based on it.
[6]These experiments, using the Ontopia Knowledge Suite by the Norwegian company Ontopia have been introduced to the Topic Map community in (in Japanese).

addition to being accessible through the web application, the texts produced in this project[7] will be made available in their raw TEI-encoded XML form as described in the following section. The other data files produced, such as the authority file for persons, which have been produced using the MADS format[8]; this is used within the Web application, but it is planned to be made available for download.

Text and markup

The text used for the digitization is the punctuated edition of the Zizhi tongjian, published by Zhonghua shuju in 1956. The text was chosen because the aim of this project was to capture structural and contextual information as far as possible, so a modern, punctuated edition seemed to be preferable. The markup applied to the text can be divided into three main categories: structural, semantic and documentary.

Structural markup

Structural markup makes the structure of the document explicit. In the case of the Zizhi tongjian, the text is divided firstly into 16 dynasties or kingdoms. Only one of these, the Tangji 唐記 in 81 juan, which is by far the most voluminous section has been the focus of the current project. Within this section, there are further subdivisions by ruler, then by era and finally by year. The editors choose to treat a calendar year as the basic unit, although obviously some changes of emperor or era do not fall at the beginning of the year, the structure is therefore not strictly nesting. Within a year, the narrative is divided into individual paragraphs. These paragraphs can already be infered from earlier texts, where white space between episodes sets them apart. In general, the editors of the Zhonghua shuju edition followed these earlier divisions, which might in fact go back to Sima Guang himself and attached numbers to them, which run sequentially within a year, however there are cases where such a paragraph is further divided, in which case the paragraphs are not numbered. The paragraphs form the most basic unit of the narration, which can be considered as reporting one event.

To record this structural division, the generic element has been used down from the Tang Records, to the eras and the years within one era. While the Zhonghua shuju edition does not make further divisions within one year, instead applying the era name of the longer part of the year to the whole year, but noting the fact that there is some change in an interlinear note, this electronic edition does mark the change where it occurs, some years are thus divided into two or more division. The change of the ruler, which also brought a change in the era, however has not been

[7]At this moment, that is the *Tangji* section of the *Zizhi tongjian* and about two thirds of the Chuxue ji 初學記.
[8]MADS is the Metadata Authority Description Schema developed by the Library of Congress.

marked with structural elements, since sometimes the narrative is proceeding in a way that allows no clean division. The following example shows the beginning of the era Wude 武德 which is slightly unusual, since it is the only one that starts at the beginning of the year rather than at the point where the era changes.

```
<div level="1" xml:id="zztj-n01" n="武德">
 <head n="武德">武德</head>
 <div level="2" type="annual" n="0618" xml:id="zztj-n01-y01">
   <head n="武德-1">
      <date n="武德-1">武德元年<note place="inline">(戊寅、六一八)</note>
               <note place="inline">
            <date>是年五月</date>受<dyn key="ch174">隋</dyn>禪，始改元。</note>
      </date>
   </head>
   <div level="3" n="1">
      <p xml:id="zztj7-6030-p01">
```

As can be seen, some additional information has been placed in the attributes to make it easier to process the texts; 'xml:id' is assigning a unique identifier to the specific element, which can be used to link to this element.

Within a paragraph (marked <p>), sentences (<s>) and phrases within sentences (<seg>) are marked, all with a xml:id attribute to uniquely identify them. Interlinear notes have been moved to the end of a paragraph and marked with <note>, the original point of attachment is given with the <anchor> element. The content of the note is further analysed; citations with attribution are marked with a <cit> element containing the attribution (<bibl>, which is not necessarily a proper textual reference) and the attributed quotation <q>, as in the following example:

```
<div level="3" n="3">
 <p xml:id="zztj7-6031-p03">
  <s xml:id="zztj7-6031-p03-s1">
   <seg xml:id="zztj7-6031-p03-s1-seg1"><date type="day" value="0627-02-06">己亥
</date>, </seg>
   <seg xml:id="zztj7-6031-p03-s1-seg2">制：</seg>
   <seg xml:id="zztj7-6031-p03-s1-seg3">「自今中書、門下及三品以上入閤議事，</seg>
   <seg xml:id="zztj7-6031-p03-s1-seg4">皆命諫官隨之，</seg>
   <seg xml:id="zztj7-6031-p03-s1-seg5">有失輒諫。」 <anchor type="note"
      xml:id="ref-zztj7-6031-p03-n.1"/></seg>
  </s>
  <note place="inline" xml:id="zztj7-6031-p03-n.1" target="ref-zztj7-6031-p03-
n.1">
   <cit xml:id="zztj7-6031-p03-c.1.1">
    <bibl> <rm key="r06728">程大昌</rm>曰：</bibl>
     <q> <dyn key="ch100">唐</dyn><dm key="dm07294">西<g type="org"
      rend="內" ref="#u20839">內</g></dm>
     <dm key="dm01809">太極殿</dm>，即朔望受朝之所，蓋正殿也。(...)
   </cit>
  </note>
 </p>
</div>
```

Semantic markup

In addition to the marking of the structure, markup is employed to distinguish semantically interesting parts in the document. The following have been marked:

- names of persons (<rm>)
- names of places (<dm>)
- names of works (<title>)
- dates, era names etc. (<y>, <date>)
- names of tribes or ethnicities (<ym>)
- names of dynasties (<dyn>)

For some of these, examples of usage can be found above. Of these items, the first two have been further analysed. In the case of personal names, attempts have been made to identify the person referred to and record this identification of the person through a 'key' attribute, which points to a unique identifier for each person. It is thus possible not only to search for persons by the names used for them (which might change or vary according to circumstances), but also find all occurrences where a person is mentioned in the text by using this identifying key. In the same way, key attributes have also been added to place names, but for the moment they have not yet been further rationalized. Dates have been analysed and months and days have been given a 'value' attribute that contains the corresponding date in the Julian calendar, for months this is the date of the first day of the month. Dynasties also have an identifying key.

Documentary markup

The Zhonghua shuju edition notes important differences or emendations in the received text, as seen in the editions used. These differences have not been verified with the source texts, but they are nevertheless documented. Like the notes, they are moved to the end of the paragraph so as to not disturb the text flow, and expressed in machine readable form as follows: The whole expressions is wrapped into a <app>(apparatus) element, which contains as its lemma (<lem>) the text as used by the Zhonghua shuju editors (but left out if there is no corresponding text) and as its reading (<rdg>) the alternative text as indicated in the modern text, its 'wit' attribute gives the witnesses that bear this variation. The 'resp' attribute on <app> has been used to record a statement of responsibility, if such one was given.

```
<p>
 <s>(...)
  <seg xml:id="zztj7-6066-p12a-s1-seg5">
   <dm key="dm08619">靈州</dm>大都督<rm key="r03311">薛萬徹</rm>爲 <dm key="dm03601
">暢武道</dm>
   行軍總管， </seg>
  <seg xml:id="zztj7-6066-p12a-s1-seg6">衆合十餘萬，</seg>
  <seg xml:id="zztj7-6066-p12a-s1-seg7">皆受 <rm key="r01299">李
   <appSpan xml:id="beg72id303625">勣</appSpan></rm>節度，</seg>
  <seg xml:id="zztj7-6066-p12a-s1-seg8">分道出<g type="org" rend="擊"
   ref="#u25802">擊</g> <ym>突厥</ym>。</seg>
  </s>(...)
  <app from="beg72id303625" resp="章">
   <lem>勣</lem>
   <rdg wit="十二行本 乙十一行本">靖</rdg>
  </app>
```

</p>

Normalization

The characters used in the text do not always reflect the specific shape that is currently used for such a character in Japan. In accordance with the textual model developed and employed here, this has been normalized to one standard form, while the specific form used in the text is recorded. Where several similar characters exist in Unicode, the one most common in Japan (except for modern simplified characters) has been used. For example, the character de 德 is frequently seen as 徳, similarly 擊 is often written as 撃. While the choice of which one to use is somehow arbitrary, the one that is more easily entered into computer systems in Japan has been considered the standard form and character usage in the text has been normalized to use this standard form. However, the character that comes closer to the one used in the printed edition is recorded and can be used for display if so desired. The <g> element is used for this purpose as seen in the example above.

Other data files

In addition to the primary text, auxiliary information that has been developed about the persons mentioned in the text. An authority file, containing the names, including variant names, of almost 30000 persons mentioned in sources on the Tang history has been created and encoded using the international standard for authority files developed by the Library of Congress, U.S.A.

A knowledge base of Tang persons

In contrast to the *Resources for Tang Studies*, which has been built in a bottom-up way by marking names of persons in the text and then extracting this information, in the *Knowledge base of Tang persons* (Pers-DB for short) information has been collected from existing reference works, anthologies like the Quan Tang shi 全唐詩 or the Quan Tang wen 全唐文, building on the series of reference works for Tang Studies Tang Civilization Reference Series 唐代研究のしおり compiled by HIRAOKA Takeo 平岡武夫 and his collaborators in the 1950s and early 1960s at our Institute, as well as more recent reference works like the Zhongguo wenxuejia dacidian 中国文学家大辞典 (唐五代卷) compiled by Zhou Zu 周祖 and his team (Beijing, Zhonghua Shuju, 1992). The aim was to include and expand on the results of centuries of research on Tang Civilization.

A technical prerequisite for the work done here was that all Unicode characters, including those in Extension B, could be represented. At the time this project was started, this limited the choices of available software considerably. To keep things simple and flexible, it was therefore decided to start out with a simple text-file based input form, which could be gradually extended and improved

on. The modelling was done by grouping related information items together in a hierarchical order (the appendix has a sample with the data for Han Yu 韓愈 filled in):

- Names
- Dates
- Places of origin
- Family relations
- Examinations
- Offices
- Works
- Biographical sources

The text files with these input records where kept in a version management system on the project server and the team would update these files when new information was found.

To make this information available on the Web, it was at a later point converted to a XML representation, which allowed more fine-grained modelling of the data. A sample of the same record in XML is also given in the appendix.

Bringing it all together

Having these rather different attempts to represent essentially the same information cries for a unification and consolidation of the data. Such a process is under way now, which presents also an opportunity to learn from the experiences and make necessary adjustments. The web site at http://tkb.mysds.jp already shows first results of this effort.

Modelling of time

One area which has not been sufficiently addressed so far is the modelling of time, which is notoriously difficult when dealing with Chinese sources, due to the idiosyncratic way of expressing units of time, but also more generally in historical material, due to the inherent of fuzziness and incompleteness of the data. Computers require data to work with and working with computerized data seems to inevitably create the illusion of precision even if no such precision exists. In attempting to model time here, a deliberate attempt has been made to allow both precision and certainty to be expressed in a way that makes these available as dimensions of the data to be modelled.

In the new model, time is made a first order object with the following characteristics:

A point in time is represented as instance of a `Date` object. This is represents a point in time with a precision of 1 day at best, but it can also be much less precise. It might contain a link to the month, which in turn links to a hierarchy of year, reign, emperor and dynasty (all of which are represented by separate database tables with the necessary entries), and it will also contain a field that gives the representation in the Julian calendar, its representation in the cyclical *ganzhi* representation. The point can also be defined by giving a range deliminated by a start point (`not_before`) and an end point (`not_after`) or a designation such a 'Mid-Tang', 'At the beginning of the reign of Xuan Zong' etc., and for all these a value for certainty can be asserted, which currently is only one of the three CERTAIN, DOUBTFUL and NEED_CHECK. This will make it possible even for date expressions not directly bearing a numeric value to have such a value assigned to them for computational purposes, thus allowing such dates also to be included in calculations, tabulations and the like, while at the same time retaining the original value of the date, which might have other connations beyond those available to computation.

To express dates in such a way introduces a certain amount of overhead, which is almost neglectible with just a few thousand records, it remains to be seen how this will scale as the collection of data grows. At the moment, the underlying calendar has only made available for the Tang period, but further preparations for other periods of Chinese history are under way.

Appendix

Example of text format for collecting and inputting information:

```
見出し：韓愈
生卒/生年：７６８
生卒/卒年：８２５
生卒/享年：
女性（性別）：
名前/姓：韓
名前/諱：韓　愈
名前/字：韓　退之
名前/排行：韓　十八
名前/謚：韓　文
名前/廟号：
名前/名号：韓昌黎　　韓吏部 韓文公
出身/貫籍：河陽　　#鄧州/ 南陽
出身/出自：昌黎
出身/寓居：
出身/生地：
出身/卒地：
家系/祖：
家系/父：韓仲卿
家系/祖先：
家系/子孫：
家系/家族：韓會（兄）
家系/姻戚：
経歴/科挙：進士@0＄貞元８
```

經歷/官職：宣武軍節度觀察推官@**拜**命＄貞元１２年７月　　＊依武寧軍節度使@**拜**命＄貞元１５　　四門博士@**拜**命＄貞元１８　　監察御史@**拜**命＄次年｜貞元１９　　陽山令@**拜**命　江陵府法曹參軍@**拜**命＄憲宗卽位 國子博士分司東都@**拜**命＄元和１年６月　　都官員外**郎**@**拜**命：仍守東都省＄元和４　　河南令@**拜**命＄翌年｜元和５　職方員外**郎**@**拜**命＄元和６　　太學博士@**拜**命＄元和７　　比部**郎**中@**拜**命＄元和８＆史館修撰@**拜**命＄元和８　　考功**郎**中@**拜**命＄次年｜元和９＆修撰@**拜**命：如故＄次年｜元和９　　（兼）知制誥@**拜**命＄元和９年１２月　中書**舍**人@**拜**命＄元和１１年１月　　太子右庶子@**拜**命 行軍司馬@**拜**命：隨裴度宣慰淮西＄元和１２年８月　　刑部侍**郎**@**拜**命 潮州刺史@**拜**命＄元和１４　　袁州刺史@**拜**命 國子祭酒@**拜**命＄元和１５年９月　　兵部侍**郎**@**拜**命＄長慶１年７月 吏部侍**郎**@**拜**命　京兆尹@**拜**命＄長慶３＆（兼）御史大夫@**拜**命＄長慶３ 兵部侍**郎**@**拜**命＄長慶３年１０月　　吏部侍**郎**@**拜**命
經歷/著作：《韓愈集》《世彩堂昌黎陝西集注・外集・遺文》　《韓昌黎文集校注》　《韓昌黎詩繫年集釋》　《順宗實錄》　《注論語》　《論語筆解》　#《全唐文・卷５４７−５６８》　#《全唐詩・卷３３６−３４５》 #《全唐詩補編・續補遺・卷５》　#《續拾・卷２４》
經歷/記事：李翺《贈禮部**尚**書韓公行狀》　《舊唐書・卷１６０・本傳》　《新唐書・卷１７６・本傳》　《唐詩紀事・卷３４》　《唐才子傳校箋・卷５》　洪興祖《韓子年譜》
備考：韓柳

Example of input data in XML

```
<個人 id="7-336">
  <見出し>韓愈</見出し>
  <姓>韓</姓>
  <諱>韓　愈</諱>
  <字>韓　退之</字>
  <排行>韓　十八</排行>
  <諡>韓　文</諡>
  <名号>韓昌黎</名号><名号>韓吏部</名号><名号>韓文公</名号>
  <生年>768</生年>
  <卒年>825</卒年>
  <貫籍>河陽</貫籍><貫籍 att="誤作">鄧州/ 南陽</貫籍>
  <出自>昌黎</出自>
  <父>韓仲卿</父>
  <家族 att="兄">韓會</家族>
  <科挙><應試><科目>進士</科目><種別>登第</種別><時>貞元 8</時></應試></科挙>
  <官歴>
    <官職><官職名>宣武軍節度觀察推官</官職名><拝命><時>貞元 12 年 7 月</時></拝命
></官職>
    <官職><官職名 name="false">依武寧軍節度使</官職名><拝命><時>貞元 15</時></拝命
></官職>
    <官職><官職名>四門博士</官職名><拝命><時>貞元 18</時></拝命></官職>
    <官職><官職名>監察御史</官職名><拝命><時>次年｜貞元 19</時></拝命></官職>
    <官職><官職名>陽山令</官職名><拝命/></官職>
    <官職><官職名>江陵府法曹參軍</官職名><拝命><時>憲宗卽位</時></拝命></官職>
    <官職><官職名>國子博士分司東都</官職名><拝命><時>元和 1 年 6 月</時></拝命></
官職>
    <官職><官職名>都官員外郎</官職名><拝命 description="仍守東都省"><時>元和 4</時
></拝命></官職>
    <官職><官職名>河南令</官職名><拝命><時>翌年｜元和 5</時></拝命></官職>
    <官職><官職名>職方員外郎</官職名><拝命><時>元和 6</時></拝命></官職>
    <官職><官職名>太學博士</官職名><拝命><時>元和 7</時></拝命></官職>
    <官職 kenkan="1"><官職名>比部郎中</官職名><拝命><時>元和 8</時></拝命></官職>
    <官職 kenkan="1"><官職名>史館修撰</官職名><拝命><時>元和 8</時></拝命></官職>
    <官職 kenkan="2"><官職名>考功郎中</官職名><拝命><時>次年｜元和 9</時></拝命
></官職>
```

<官職 kenkan="2"><官職名>修撰</官職名><拜命 description="如故"><時>次年｜元和9</時></拜命></官職>
<官職><官職名 type="兼">知制誥</官職名><拜命><時>元和 9 年 12 月</時></拜命></官職>
<官職><官職名>中書舍人</官職名><拜命><時>元和 11 年 1 月</時></拜命></官職>
<官職><官職名>太子右庶子</官職名><拜命/></官職>
<官職><官職名>行軍司馬</官職名><拜命 description="隨裴度宣慰淮西"><時>元和 12 年 8 月</時></拜命></官職>
<官職><官職名>刑部侍郎</官職名><拜命/></官職>
<官職><官職名>潮州刺史</官職名><拜命><時>元和 14</時></拜命></官職>
<官職><官職名>袁州刺史</官職名><拜命/></官職>
<官職><官職名>國子祭酒</官職名><拜命><時>元和 15 年 9 月</時></拜命></官職>
<官職><官職名>兵部侍郎</官職名><拜命><時>長慶 1 年 7 月</時></拜命></官職>
<官職><官職名>吏部侍郎</官職名><拜命/></官職>
<官職 kenkan="3"><官職名>京兆尹</官職名><拜命><時>長慶 3</時></拜命></官職>
<官職 kenkan="3"><官職名 type="兼">御史大夫</官職名><拜命><時>長慶 3</時></拜命></官職>
<官職><官職名>兵部侍郎</官職名><拜命><時>長慶 3 年 10 月</時></拜命></官職>
<官職><官職名>吏部侍郎</官職名><拜命/></官職>
</官歷>
<著作><書物>《韓愈集》</書物><書物>《世彩堂昌黎陝西集注・外集・遺文》</書物><書物>《韓昌黎文集校注》</書物><書物>《韓昌黎詩繫年集釋》</書物><書物>《順宗實錄》</書物><書物>《注論語》</書物><書物>《論語筆解》</書物><書物>#《全唐文・卷547—568》</書物><書物>#《全唐詩・卷 336—345》</書物><書物>#《全唐詩補編・續補遺・卷 5》</書物><書物>#《續拾・卷 24》</書物></著作>
<記事><書物>李翺《贈禮部尚書韓公行狀》</書物><書物>《舊唐書・卷 160・本傳》</書物><書物>《新唐書・卷 176・本傳》</書物><書物>《唐詩紀事・卷 34》</書物><書物>《唐才子傳校箋・卷 5》</書物><書物>洪興祖《韓子年譜》</書物></記事>
<備考>韓柳</備考>
</個人>

Bibliography

Berners-Lee, Tim(1999), *Weaving the Web*. Harper San Francisco:

Berners-Lee, Tim; **Hendler, James**; **Lassila, Ora** (2001), *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, in: *Scientific American*. URL: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Berners-Lee, Tim; **Shadbolt, Nigel**; **Hall, Wendy** (2006), *The Semantic Web Revisited*, in: *IEEE Intelligent Systems* Vol.21, p. 96-101. URL: http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf

TEI Consortium, (2007), *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. URL: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html

Biezunski, M.; **Bryan, M.**; **Newcomb, S.**(2002), *ISO/IEC 13250, Topic Maps (Second Edition)*. URL: http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0322.htm

Boag, Scott; **Chamberlin, Don**; **Fernandéz, Mary F.**; **Florescu, Daniela**; **Robie, Jonathan**; **Siméon, Jérôme** (2007), *XQuery 1.0: An XML Query Language*. URL: http://www.w3.org/TR/xquery/

Library of Congress, (2004), *Metadata Authority Description Schema (MADS)*. URL: http://www.loc.gov/standards/mads/

Naitoh Motomu 内藤求; (2006), トピックマップ入門. 東京電機大学出版局.

O'Reilly, Tim (2005), *What Is Web 2.0*. URL: http://www.oreillynet.com/lpt/a/6228

Voss, Jakob (2007), *Tagging, Folksonomy & Co - Renaissance of Manual Indexing?*, in: *cs/0701072*, (preprint, 12pages) -. URL: http://arxiv.org/abs/cs/0701072

Wittern, Christian ; **Akiyama Yōichirō** 秋山陽一郎, (2004), 唐代ナリッジベースに向かって, in: 漢字文献情報処理研究 Vol.5, p. 20-24.

Wittern, Christian (2005), *The Text in the Age of Digital Reproduction*. Chung-Hwa Buddhist Journal, Taipei, Taiwan.

Wittern, Christian; **Nakadate Hamana**; 中楯はまな (2007), 漢讀：新しいテキスト・モデルに基づいた東洋学文献研究支援ツール, in: 情報処理学会 研究報告 Vol., p. 9-16.

Wittern, Christian (2007), *Digital Text, Meaning and the World : Preliminary considerations for a Knowledgebase of Oriental Studies*. Institute for Research in the Humanities: Kyoto, Japan.