

# The Chan School as seen through the Knowledgebase of Tang Civilization

Christian Wittern

## 1. Introduction

The title of this presentation might suggest that something called 'Knowledgebase of Tang Civilization' exists and something called 'Chan school' will be analyzed within. Unfortunately, this is not (yet) the fact. Work on the Knowledgebase has just begun and instead of the Chan school as a whole, there is so far only one text, the 祖堂集 *Zutang ji*, that will be considered.

What this paper in fact will be trying to do is to use elements, building blocks, that will become available as part of the Knowledgebase and try to play around with them to see how they could be used and what kind of questions could be answered using these things. The main purpose of this undertaking will not yet be to produce new insights to the understanding of Tang Civilization or the Chan school (although I hope that this will also be the case), but rather to see what needs to be taken into consideration for the construction of the Knowledgebase, where the current plan might need reconsideration or a shift of focus. So rather than waiting for the whole construction of the house to be finished before moving in and then to discover that all the electric outlets have been misplaced and the ceiling in some rooms has been forgotten, I will imagine to move in and live within the Knowledgebase, although not even the walls have been finished yet, let alone the roof.

## 2. The Knowledgebase of Tang Civilization

A few years ago, the prospect of having access to a large amount of digitized data promised to give a completely new direction to the field of Chinese Studies. Although today we have databases such as the 四庫全書, as well as many other texts, some of them even freely available on the Internet, the benefits of this has been limited. There are many reasons for this, not all of them technical. Of the technical reasons, the limited, idiosyncratic interface that each database provides, and the unstructured data it operates on are probably the most important ones.

The Knowledgebase, which is currently under development at the Institute for Research in Humanities<sup>1</sup>, is an attempt to remedy this situation by providing 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 and generally support research related to Tang Civilization. It's main

<sup>1</sup> More information about the methodology and working plan have been given in and as well as on the website <http://tkb.mydns.jp>.

point of access for researchers will be a web application, but other interfaces are also envisioned.

Conceptually, the Knowledgebase is divided in a 'information layer' and a 'resource layer.' The resource layer will hold mostly texts, but other types of information, such as maps, images etc. might also be added. The information layer will hold secondary information about entities that exist in the resource layer, but will not be restricted to that. So a person mentioned in a text, for example, might also have an entry in the information layer, which will provide background information on his biography, career, family, works and so on. These two layers are interrelated, going back and forth between them will be seamlessly possible.

The information layer can be used to focus on specific elements or parts of the resource layer. For example, it would be possible to request a list of persons, who held the post of prime minister (zaixiang, 宰相) during the 8th century and analyze i.e. their places of origin, or their period in office. In the future, one could also search in the extant works of these prime ministers, possibly limiting the search to a certain area or topic.<sup>2</sup>

Methodologically, the construction of the Knowledgebase does rely foremost on primary sources; these will be analysed, structured, integrated to the resource layer and some information will be extracted to the information layer. Although secondary sources and reference works are also employed in that process, it is a principle that every piece of information in the Knowledgebase will be labelled in a way that its origin and scope can be traced. To lay a foundation for specialized information, which will be encoded in due time, historiographic records have been selected to provide the most basic information about the events, the persons involved therein and the time and place of such happenings, administrative and geographic structure and so forth; the start has been made with the orthodox historiographical records 舊唐書 *Jiu Tang shu* by 劉昫 Liu Xu and others (945), 新唐書 *Xin Tang shu* by 歐陽修 Ouyang Xiu and others (1060) and 資治通鑑 *Zizhi tongjian* 司馬光 Sima Guang (1084)<sup>3</sup>. It goes without saying, that these sources are not first hand historical sources, but rather compilations and selections made from at some point in time after the events reported therein, from a certain perspective. Nevertheless, the Knowledgebase will represent the information derived from these sources at face value with source attribution. In due time, as more information finds its way into the Knowledgebase, there will be different perspectives and other sources, both primary and secondary, that will provide the necessary corrective.

<sup>2</sup> It should be made clear here, that most likely there will not be a large corpus of full texts available from the beginning. The Knowledgebase will however provide the technical means to do interact behind the scenes with other collections, that might hold these texts.

<sup>3</sup> These texts and more are used internally to prepare the information contained in the information layer of the Knowledgebase. Due to legal restrictions not all of them might be available in the resource layer at the time of publication of the Knowledgebase.

The information in the knowledgebase is organized along the following information axis:

- Chronologic and astronomic: calendar and time
- Toponomastic: placenames and georeferences to there locations, administrative geographical units, digital maps.
- Prosopographic and onomastic: personal names, dates and activities of people of the Tang.
- Bibliographic and artistic: works created during the Tang, including texts, artefacts and buildings
- Historiographic: events of importance and influence
- Epigraphic: images and transcriptions of epigraphic sources

These items are maintained in the information layer usually as the smallest possible atomic units. The atomic units are then connected to other units to form higher level features<sup>4</sup>. In the example above, the entry for one person would be such an atomic unit, as would be the role of 'prime minister'. To record that a person has been holding the post of prime minister at a certain period of time, the item for the person would be 'associated' to the item identifying the role; the timespan can be expressed using 'scope', which allows to specify the validity conditions for the assertion.

This methodology is straightforward in simple cases, where enough is known to assert the subject identity<sup>5</sup> of what is to be encoded as atomic units. In many cases, however, the subject identity can not easily be asserted. Two identical names encountered at two different locations might refer to the same person or not, to decide about this, more contextual information is required. The system employed here allows to provide or derive the contextual information as needed, without necessarily asserting subject identity.<sup>6</sup>

<sup>4</sup> The specific methodology employed here is based on the topic map paradigm, which uses the basic concept of 'topics', 'associations' and 'occurrences' as handles to map out the information space. Topic maps are formally defined in International Organization for Standardization, *ISO/IEC 13250, Information technology - SGML Applications - Topic Maps* Geneva, 2000. More information about topic maps can be found for example in Steve Pepper, *The TAO of Topic Maps Finding the Way in the Age of Infoglut*, in: [<http://www.ontopia.net/topicmaps/materials/tao.html>]

<sup>5</sup> Subject identity is another concept used in the topic map paradigm, it is used to make it possibly to uniquely identify the subject of a topic, independently of his name or properties. This role is similar to that of the social security number, used in some countries to identify the citizens.

<sup>6</sup> There is of course a whole philosophical debate waging around this very issue. It is important to note, that the methodology employed here does not rely on the assertion of subject identity of the objects in the Knowledgebase, but of course it can use the subject identity if such an identity is established.

These information items are organized in hierarchical ontologies<sup>7</sup>. This allows to access the information also based on their position within the hierarchy, or on the relation with other items. For geographical locations, like a city for example, such a hierarchy would consist of the upper administrative units it belongs to; for persons this could consist of the family line, but also the region of origin, the school or tradition of thought, the offices held during the career or in the case of monks also the ordination line and line of transmission.

The creation and maintenance of such ontologies is by itself an act of interpretation of the data and certainly not universally assertable. The Knowledgebase does allow multiple, conflicting ontological hierarchies coexist and the users will be able to add to or modify the ontologies.

In the following sections, some of the more important subtrees of the ontology will be discussed, as they are currently maintained in the Knowledgebase. This is very provisional work in progress.

## **2.1. Chronological Axis**

Many information items in the Knowledgebase will have a temporal aspect, which will hook into what will be called the chronological axis here. Temporal references might be very specific (e.g. Kaiyuan 3rd year, 1st day of the second month, which had the cyclical sign ), very generic (8th century) or unspecified (dates unknown). In addition to that, the framework for calculating the calendar has changed several times during the Tang dynasty, and other calendar systems, for example the western (Julian) calendar is used by scholars. All this will need to be representable in the Knowledgebase.

To solve this problem, the relations between different calendric items are described in this subtree of the ontology. As the smallest unit for calendric items the 'day' has been chosen. All days in the almost 300 years of the Tang dynasty (altogether 105544 days) have been given a unique identifier, which is the 'Julian

<sup>7</sup> Ontology, is used here as a term used in the field of knowledge representation. John F. Sowa defines it as follows: "The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalog of the types of things that are assumed to exist in a domain of interest D from the perspective of a person who uses a language L for the purpose of talking about D. The types in the ontology represent the predicates, word senses, or concept and relation types of the language L when used to discuss topics in the domain D." (*Knowledge Representation*, (ThomsonLearning 2000) p492.

Day', first proposed by Joseph Justus Scaliger<sup>8</sup> and today generally used in astronomic calculations.<sup>9</sup>

The Julian Day number will also serve as a convenient means to calculate periods of time, for example the length of reign periods of the Tang emperors. The table 1 shows a list of the Tang emperors and empress, together with the length of their period on the throne and the fraction their reign holds of the total length of the Tang.<sup>10</sup>

Table 1. Tang Emperors and their reigns

Name	First day on throne	Days on throne	Percentage of total number of days (105544)
高祖	0618-06-18	3141	2.97
太宗	0627-01-23	8416	7.97
高宗	0650-02-07	12403	11.75
中宗	0684-01-23	35	0.03
睿宗	0684-02-27	348	0.32
武則天	0685-02-09	8273	7.83
中宗	0707-10-05	1004	0.95
少帝	0710-07-05	45	0.04
睿宗	0710-08-19	755	0.71
玄宗	0712-09-12	16040	15.19
肅宗	0756-08-12	2100	1.98
代宗	0762-05-13	6483	6.14
德宗	0780-02-11	9334	8.84

<sup>8</sup> The French scholar Joseph Justus Scaliger (1540-1609) was interested in assigning a positive number to every year without having to worry about B.C.E. / C.E. He invented what is today known as the Julian Period. The Julian Period probably takes its name from the Julian calendar, although it has been claimed that it is named after Scaliger's father, the Italian scholar Julius Caesar Scaliger (1484-1558). Scaliger's Julian period starts on 1 January 4713 B.C.E. (Julian calendar) and lasts for 7980 years. 2004 is thus year 6717 in the Julian period. After 7980 years the number starts from 1 again.

<sup>9</sup> Of immense help in calculating these dates and providing the base for the ontology was the program "WHEN", written by suchowan@日経mix, available at his website <http://www.asahi-net.or.jp/~dd6t-sg/> and 唐の曆(T'ang Calendar) by 平岡武夫 Hiraoka Takeo (Kyoto, 1954).

<sup>10</sup> A word of warning should be inserted here about the use of figures here. These figures have been calculated for use within the framework of the Knowledgebase, as will be shown. More than the exact value, the value in its relationship to other values should be considered. In most cases, the value should be seen just as giving an estimate, not a precise figure.

Name	First day on throne	Days on throne	Percentage of total number of days (105544)
順宗	0805-09-01	146	0.13
憲宗	0806-01-25	5494	5.20
穆宗	0821-02-09	1450	1.37
敬宗	0825-01-29	774	0.73
文宗	0827-03-14	5076	4.80
武宗	0841-02-04	2193	2.07
宣宗	0847-02-06	5063	4.79
懿宗	0860-12-17	5113	4.84
僖宗	0874-12-17	5163	4.89
昭宗	0889-02-04	6695	6.34

## 2.2. Toponomastic Axis

Inspired by the possibilities of electronic maps, geographical information systems and their applications, research and implementation of toponyms, gazetteers and Chinese toponomastics in general has seen considerable activity<sup>11</sup>.

In the Knowledgebase, so far only a hierarchical list of names and their administrative subsumption has been prepared, mainly derived from the geographical treatises in *Jiu Tang shu* and *Xin Tang shu*; this list has been used for the research reported here. The record of administrative changes concerning name, area and subordination that is available to considerable detail in *Jiu Tang shu*, at least for the time until 756, and the information about population, distance to the capitals, local products etc. will be subsequently integrated.

## 2.3. Prosopographic and Onomastic Axis

Considerable effort has been made so far to collect information on people living in Tang times. Work was started with analyzing the collected biographies in the *liezhuan* sections of *Jiu Tang shu* and *Xin Tang shu*. About 3000 names of persons have been derived from there, subsequently information about the various names they used and other information like whether they took examinations, the place of origin, birth and death dates and so on have been collected. Currently,

<sup>11</sup> See for example the 'Chinese Civilization in Time and Space' [<http://ccts.ascc.net>] project coordinated by 范毅軍 Fan I-chun, at Academia Sinica in Taiwan, and the 'China Historical GIS' [<http://www.people.fas.harvard.edu/~chgis/>] coordinated by a team of scholars at Harvard University (USA) and Fudan University (PR China) under Peter K. Bol. There has also been some activity coordinated by Lewis Lancaster as part of the 'Electronic Cultural Atlas Initiative' [<http://ecai.org>] project.

information derived from the *Zizhi tongjian* is added. Obviously, it will take quite some time to even digest the current contents and then fill in the information on persons missing. The goal is, however, to have an entry for every person that is known to have lived in Tang China and to collect as much information and references about this person as possible. Since this work is very much in progress, any results reported here involving these data should be taken as preliminary and might change as new data become available.

### 3. Estimation of completeness of the historical records

In the texts that are so far in the Knowledgebase, that is *Zizhi tongjian*, *Xin Tang shu* and *Jiu Tang shu*, among other things all occurrences of reign names have been marked. This is necessary, in order to distinguish reign names from other proper names used in the text, especially where there is disambiguity, for example in the case of 長安 chang'an, which was the name for the reign period from 701 to 705, but is also used as a place name.

With this information, a simple first estimate of the completeness of coverness of the historical records can be tried<sup>12</sup>. For this purpose, we assume that in reasonably complete records, the frequency of distribution of reign names should roughly match the length of the reign weighted against the total length of the dynasty. This estimate will of course be distorted by records of some dramatic events that might require more extensive coverage, but it should still be good enough as a first estimate, especially if only used to check for the existence of suspicious undercoverage.

The texts under discussion here are of different types. The first part, 本紀 Benji of *Jiu Tang shu* and *Xin Tang shu* as well as the *Zizhi tongjian* are chronologically arranged, so the length of sections (in number of characters) for the entries of years and reign periods can be easily calculated. The result of this, weighted against the percentage this period against the whole length of the Tang is given in Figure 1. For the biographical 列傳 Liezhuan section, this is not possible, here the number of occurrences of the different periods have been calculated. This should give a better measure of the overall coverage, since in chronological records, gaps are easily visible and could be covered up, whereas in the biographical sections, the extent of coverage is not easily detectable. The result of this second calculation is given in Figure 2. As can be seen, the chronological works thrive for a certain degree of balancedness, maybe with the exception of the *Zizhi tongjian*, where in the second half of Tang, the record seems to get more scattered. However, it is

<sup>12</sup> At several occasions parts of the historiographic records have been lost during the Tang, probably the most severe loss was endured during the luting of the capital city Chang'an by An Lushan in the 6th month of 756. See on this topic Denis Twitchett, *The writing of Official History under the T'ang* p16 and 46 and Robert des Rotours, *Histoire de Ngan Lou-chan (Ngan Lou-chan Che tsi)*, p273n4. Before other interpretative attempts, an estimate of the completeness seems to be necessary.

clearly visible that in the biographical records more persons from the first half of the dynasty are included.

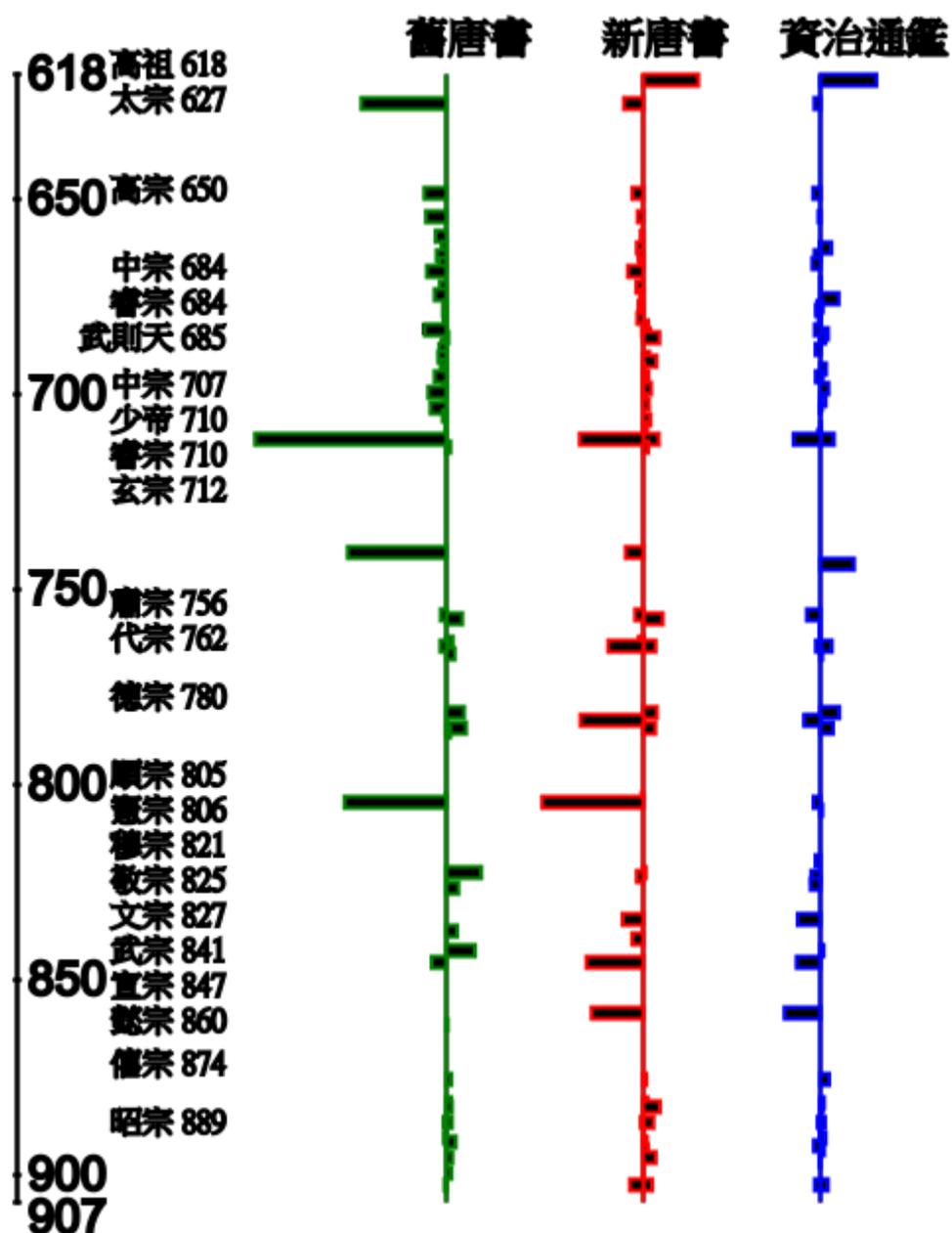


Figure 1. Relative length of the Benji annals and the Zizhi tongjian to the length of the reign periods<sup>13</sup>

<sup>13</sup> The figure shows the relative length of the period calculated as the number of characters subtracted from the relative length of the period in days in the right half, a scale of years and the names of the Tang emperors in the left half. Negative values are drawn to the left of the center line, positive values to the right. Values to the left side thus shows a relative scarcity of records, while positive values, extending to the right, show relative abundance of records. While longer

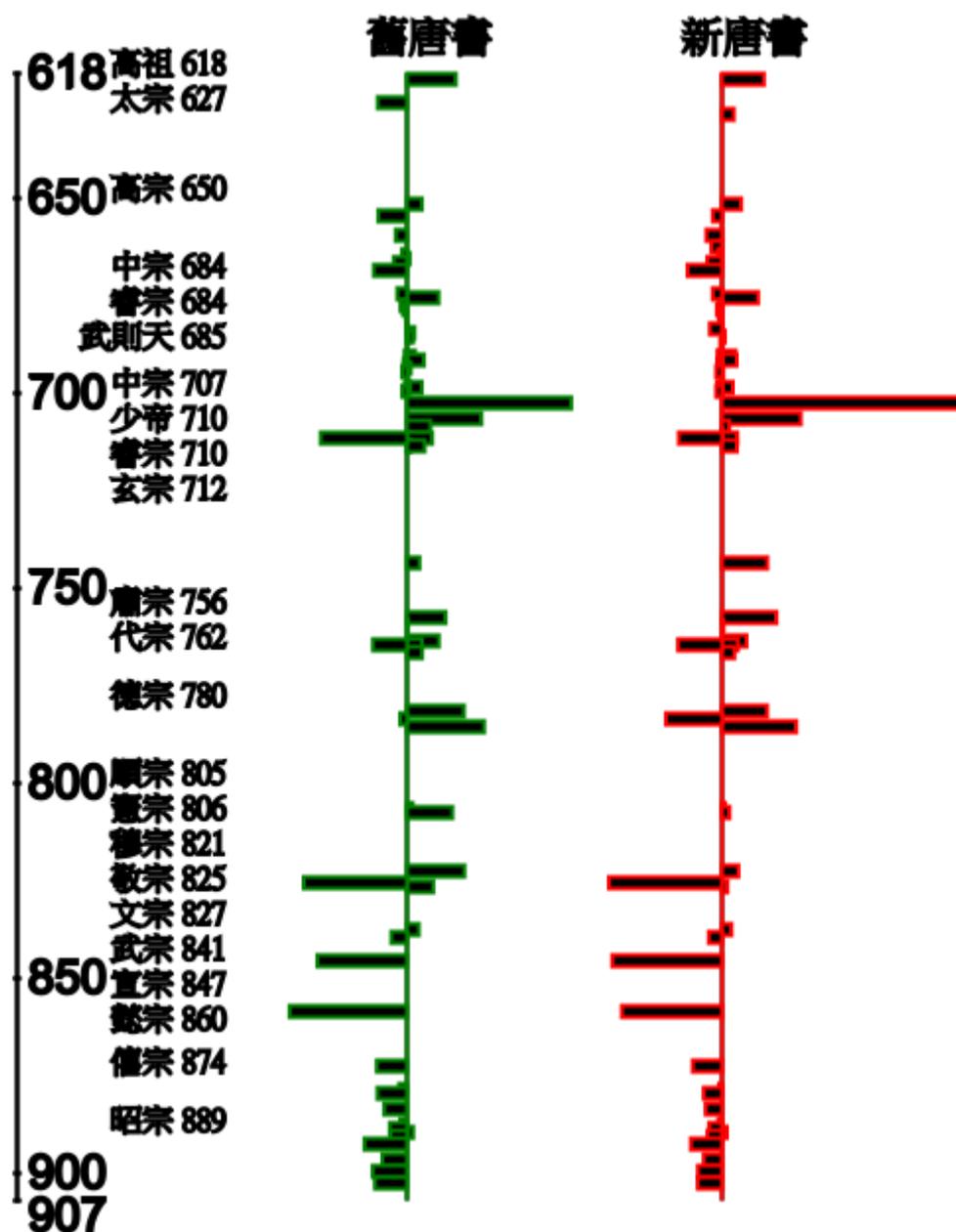


Figure 2. Relative length of the Liezhuan sections in Jiu Tang shu and Xin Tang shu<sup>14</sup>

eras show some distortions, the relative scarceness of records for the Kaiyuan period of Xuan Zongs reign (713-742) is clearly visible.

<sup>14</sup> The calculation is done in a similar way as the previous figure, in this case the number of matches for each area within the biographical section has been compared to the overall number of its, which is used to measure the relative coverage of the reign periods. A relative scarcity in the latter half of the Tang, especially in the 9th century is clearly visible.

## 4. A glimpse of the Chan school through the *Zutang ji*

As described above, the ‘Tang Knowledgebase’ is so far composed only of material that belongs to the genre of official historiographic literature. To find out, how the Chan school is represented in these works, this will be contrasted with findings from the analysis of a chronic of the Chan school, the *Zutang ji Zutang ji*. While this is not a historiographical record, the text does take the pose of telling the history of the Chan school and attempts to create its view of how the Chan school came into being. It was compiled by the Chan monks 靜 Jing (no dates) and 均 Yun (no dates), disciples of Chan master 文燈 Wendeng (884-972) of the 招慶寺 Zhaoqing monastery in 泉州 Quanzhou in 952. It contains records of 246 individual Chan masters, although not all of them are historical: the work starts with the seven Buddhas of the past, and then works down the traditional lineage of the Chan school, which was already well established at the time this work was compiled. The bulk of the entries, however is about Chan masters active during the Tang.

The text of the *Zutang ji* was discovered at the beginning of the last century among the more than 80000 wooden printing plates of the Korean Tripitaka, which was carved and printed in the middle of the 13th century and is now held at the temple Haiensa in Korea. While written in Quanzhou, in the south of today's Fujian province in China, the work was lost in China and unknown until its rediscovery. It has gathered considerable interest since then, especially by the pioneer of modern Chan studies, Yanagida Seizan, who published an index to this work in the 1980s. In the 20 years since that publication, a considerable number of studies have been published, focusing both on the text as an early witness of the history of Chan, and on its value for research on the history of the early written vernacular of Chinese. One more index of the whole work, and four typeset editions have been published so far.<sup>15</sup>

Since the text is only transmitted in one single woodblock rendering, there are a number of ideosyncratic character usages, that provide considerable difficulties in creating an electronic edition for this text. The present author has worked on such an edition for a number of years, but it has not been readied for publication yet. For the purpose of further analyzing the content in a manner similar to the texts mentioned above, some textual features have been marked, such as names of persons, places and institutions, dates, and titles of books.

In the same way as the texts above, the *Zutang ji* has been submitted to an estimation of coverage based on the occurrence of reign names (See 3). Since this

<sup>15</sup> Most recently, there is the work done by Christoph Anderl, in his forthcoming PhD thesis *Studies in the Language of Zu-tang ji* (Univ. of Oslo), who is also preparing a complete translation of the work into English. 古賀英彦 Koga Hidehiko has published a rendering in traditional Japanese with notes as 訓注祖堂集 *Kunchū Sodōshū* as Vol. 8 of the 花園大学国際禅学研究所研究報告第八冊 *Studies of the International Research Institute for Zen Buddhism*, 2003.

text is of a different genre, the overall occurrence of reign names is rather small, so the estimate has to be treated with even more caution, but nevertheless quite contrary to the biographical sections in 2 above, to which this genre is closest, the 3 shows more frequent occurrences in the latter half of the Tang. This should come as no surprise however, since the development of the Chan school has taken place in that period, it should rather be seen as a confirmation that this kind of analysis does indeed provide significant data.

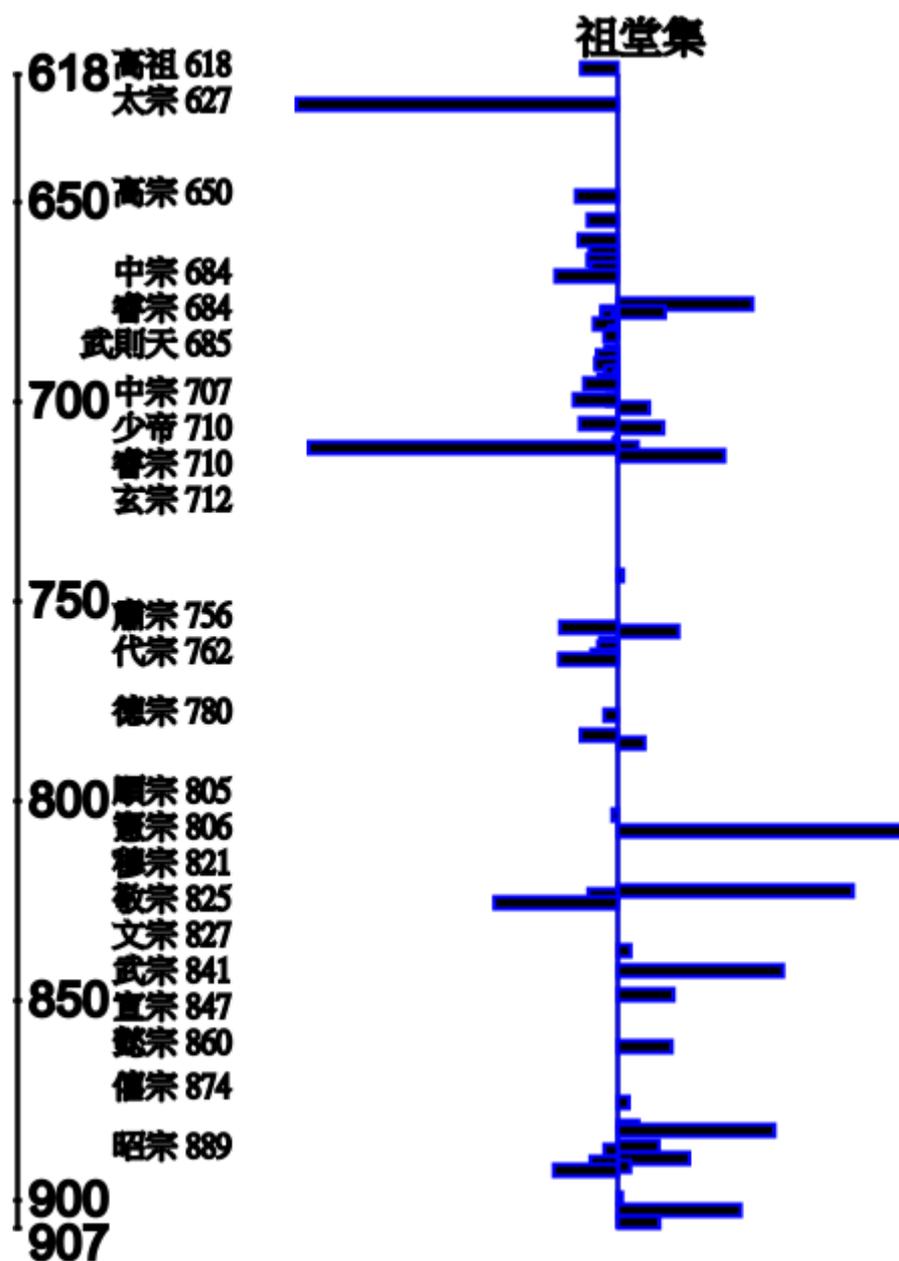


Figure 3. Relative length of the entries of the Zutang ji<sup>16</sup>

Far more interesting is of course not to compare only the coverage, but also other aspects of the historiographical records and the Chan chronicle. During the Tang period, there were of course a number of high officials with strong relations to members of the Chan school, who do have a biography in the dynastic histories.

There is for example 裴休 Pei Xiu (797-870)<sup>17</sup>, a high official, who even held the post of prime minister for some time and a lay follower of 黃檗希運 Huangbo Xiyun (died ca. 850)<sup>18</sup> who even wrote a preface to an edition of his records, is mentioned in *Zutangji* (2-043-08, in the entry on 磁州法如 Cizhou Faru and 4-136-09 in the entry on Huangbo), but neither in his biography in the *Jiu Tang shu* (14:4593-4) nor in the *Xin Tang shu* (17:5371-2) is there any trace of his interest for Chan, and no name is mentioned that would also appear in the *Zutangji*.

Given this fact, there is little hope in finding names mentioned reciprocally in both categories of texts and indeed an analysis of the cross-section of the list of personal names in both texts gives only very few names<sup>19</sup>. The only exception is 神秀 Shenxiu (?-706), who does have an entry in the *Jiu Tang shu* (16:5109-11) in the section on healer, geomantics, traumathurgs and necromants, which also includes Daoists and Buddhist monks, and in his biography, 弘忍 Hongren (601-74), 慧能 Huineng (638-713), 普寂 Puji (651-739) and 義福 Yifu (658-736) are mentioned in passing.

Other high officials mentioned in the *Zutangji* are 韓愈 Han Yu (768-824), who is mentioned in the entry of 大顛寶通 Dazhen Baotong (732-824) (ZTJ 2-001-05), engaged in a discussion on the reality of the power of Buddhist relics<sup>20</sup>, and 李翱 Li Ao (772?-841?), who is participating in an exchange with 藥山惟儼 Yaoshan Weiyuan (745-828) (ZTJ 1-169-09).<sup>21</sup>

A similar cross examination has been started for place names, but since the identification of place names in the information layer has only just begun, the results are very preliminary and I refrain from giving a table here. A cursory

<sup>16</sup> The calculation algorithm is the same as in the previous figure. Due to the relative small number of overall hits, the amplitude of the bars is larger, this should not be misunderstood as significant. There is however significant more activity recorded for the period after 800, especially if compared to Figure 2.

<sup>17</sup> On Pei Xiu, see the excellent biographical essay by 吉川忠夫 Yoshikawa Tadao 裴休傳 *Haikyū den*, in: 東方學報第六十四冊 *Journal of Oriental Studies* (1992), p115-277.

<sup>18</sup> In the *Jingde Chuandenglu* T51, p293a-c, there is even a record for him as a dharma heir of Huangbo.

<sup>19</sup> It has to be mentioned here, that the names are not yet completely normalized and identified, so there is a possibility that some names have been missed.

<sup>20</sup> 西脇常記 Nishiwaki Tsuneki discusses the background of this discussion in 唐代の思想と文化 *Thought and Culture in Tang China*, p265-272, especially p270.

<sup>21</sup> A number of officials are mentioned in the *Zutangji* only with the surname, like 召相公, 雷相公, 太尉相公, 于迪相公, 韋曹相公. Their identity has not been established with the current methods. For this purpose, the Knowledgebase will need to provide a way to use the contextual information as outlined above and assist in this process.

examination of this list seems to indicate that there are more places in the center and north mentioned in the historiographical records and the *Zutang ji* has a majority of places at the coast and in the south, which again is not surprising at all. A more closer analysis into the toponomastics is however required, to determine at what period the placenames mentioned in the *Zutang ji* had been used, this might reveal some more background on the formation of the text.

#### 4.1. Lineage

One other aspect deserves to be discussed here. One of the fundamentals of the self-image maintained by the Chan tradition is the concept of the direct transmission from Shakyamuni Buddha to the present day in a single lineage that has never been interrupted. This is to say, that there is exactly one master, who functions as the spiritual father is the one that transmits his understanding to his disciple and gets the seal of approval (印可 *yinke*). While this, as has often been pointed out<sup>22</sup>, is not maintainable as a historic fact, even within the very texts that are propagating this image, of which the text under discussion here is a prominent example.

The text as we have it today, does allow for a completely different line of investigation. Instead of just taking the supposed line of transmission, we might look at the actors appearing on stage in the records for the individual masters.

The analysis of these records will be conducted along the following lines: Since the occurrences of personal names have already been marked in the text of the *Zutang ji*, it is rather straightforward to derive a list that gives for each master the list of names that appear in his record. While the mere occurrence does not indicate what kind of relationship the two persons might have held, we will assume that there has been some kind of relationship, which will be called 'relation by co-occurrence' as a heuristic method, however if the name does identify either the master or a disciple, as recorded in the lineage charts, then this relation is recorded. Secondly, if a name occurs frequently, a higher degree of relation is assumed. It is also checked whether the occurrence is reciprocely, that is, if there is also a record for a name mentioned here, an occurrence of the current name in that record would indicate some kind of reciprocity in the relationship.

The result of this investigation shows the network like structure that is visible as the subtext of this chronicle, which claims on the surface to be monolinear. While a visualization of the network structure could more vividly show the way the actors in this text are interconnected with each other, a table will be used here to show the results. Only the 20 masters with the most occurrences are shown here, ordered by number of occurrences. For each of these masters, the total

<sup>22</sup> See for example most recently the discussion in John McRae, *Seeing through Zen*, Chapter 1, Looking at Lineage, p1-21.

number of names mentioned in the record, the number of matches of disciples, and reciprocal matches is given.<sup>23</sup>

Table 2. Occurrence of names in entries of the *Zutangji*

Rank	Name	Total	Within the lineage (master or disciple)	Reciprocal occurrence	Other names
1	仰山 (5-050-08)	326	滄山 (79)	雪峰 (2, 8)	達摩 (13)、慧能 (9) 道存 (8)、慧明 (8)、慧寂 (6)、道明 (6)、神會 (5)、
2	南泉 (4-107-03)	188	趙州 (16)、歸宗 (14)、逍遙 (4)、石霜 (1)	保福 (2, 5)、洞山 (2, 5)、雲岳 (10, 1)、報慈 (1, 1)	曹山 (21)、道吾 (14)、王老師 (13)、黃蘗 (12)、藥山 (8)
3	菩提達摩 (1-069-08)	186	惠可 (12)、般若多羅 (3)		武帝 (12)、馬祖 (11)、志公 (7)、神光 (6)、楊衍 (5)
4	雪峰 (2-099-01)	149	長慶 (23)、玄沙 (8)、鏡清 (3)、保福 (2)、長生 (1)	仰山 (8, 2)、德山 (4, 6)、岳頭 (3, 9)、欽山 (3, 2)、趙州 (2, 4)、鵝湖 (1, 1)	疏山 (5)、達摩 (5)、滄山 (4)、徑山 (4)
5	藥山 (1-168-04)	149	雲岳 (35)、石頭 (5)	、道吾 (33, 7) 滄山 (3, 2)、長	百丈 (8)、曹山 (3)、

<sup>23</sup> It should be noted that the research documented in this table is not yet concluded; some of the names masters (abbreviated names and other appellations) have not yet been mapped to the information layer. The table should therefore be seen as visualization of the general trend, without yet providing the necessary detail. The first column gives the rank of the entry in terms of frequency of occurrences of personal names in that entry, the total number is given in the third column. In the second column, the page reference to the text is given in parentheses. In the other columns, the numbers in parentheses indicate the number of occurrences, for the reciprocal occurrences, the second number is the value for the corresponding entry.

Rank	Name	Total	Within the lineage (master or disciple)	Reciprocal occurrence	Other names
				慶 (2, 3) 、紫玉 (1, 3)	
6	釋迦牟尼佛 (1-009-05)	142			阿難 (3) 迦葉 (17) 、阿闍世王 (7) 、慈氏 (2) 、傳教阿難 (1) 、賢阿難 (1) 、慶喜阿難 (1)
7	保福 (3-077-05)	126	雪峰 (6) 、長慶 (5) 、鼓山 (2) 、安國 (1)	南泉 (5, 2) 、岳頭 (3, 2) 、東寺 (2, 1) 、趙州 (1, 1) 、雲居 (1, 2)	招慶 (30) 、文殊 (11) 、洞山 (7) 、無著 (4) 、
8	瑞雲寺和尚 (5-113-03)	109			普賢 (45) 、文殊 (20) 、智通 (7) 、該通 (4) 、遮那 (4) 、提婆 (3) 、行通 (3) 、龍樹 (2) 、百牙 (1)
9	岳頭 (2-089-07)	109	雪峰 (18) 羅山 (11) 、	德山 (21, 2) 、保福 (2, 3)	滄山 (5) 、洞山 (4) 、長慶 (4) 、夾山 (4)
10	洞山 (2-049-12)	105	雲岳 (6) 、曹山 (3)	雲居 (9, 11) 、南泉 (5, 2) 、五洩 (4, 3) 、疏山 (3, 1)	雪峰 (11) 、顏 (7) 、西峰 (5) 、南泉 (5) 、鳳池 (4) 、
11	雲岳 (2-009-04)	103	洞山 (44) 、藥山 (8)	滄山 (9, 5) 、道吾 (5, 5) 、百丈 (3, 4) 、雲居 (2, 1) 、長慶 (1, 1) 南泉 (1, 10)	石頭 (2) 、疏山 (2)

Rank	Name	Total	Within the lineage (master or disciple)	Reciprocal occurrence	Other names
12	大迦葉 (1-027-03)	88			阿難 (29)、阿闍世王 (7)、慈氏 (2)、傳教阿難 (1)、賢阿難 (1)、慶喜阿難 (1)
13	滄山 (4-124-13)	84	仰山 (31)、百丈 (3)	雲岳 (5, 9)、道吾 (5, 7)、藥山 (2, 3)	報慈 (3)、隱峰 (3)、曹山 (2)、雲居 (2)
14	慧能 (1-089-11)	76		弘忍 (3, 1)	薛簡 (8)、道誠 (7)、印宗 (4)、雲大師 (3)、達摩 (3)、龍花 (2)、
15	馬祖 (4-033-02)	71	百丈 (9)、汾州 (4)、西堂 (2)、南泉 (1)		黃三郎 (4)、漳南 (3)、忠國師 (3)、亮座主 (3)、石門 (2)
16	禾山 (3-120-01)	68	洞山 (7)、曹山 (2)	石霜 (4, 3)	岳 (4)、興平 (3)、迦葉 (3)、九峰 (3)、雲岳 (2)、道吾 (2)
17	玄沙 (3-037-05)	65	雪峰 (8)、長生 (1)	中塔 (4, 5)	雲 (6)、中塔 (6)、靈雲 (5)、慶 (5)、志超 (2)、順德 (2)、
18	趙州 (5-037-03)	62	南泉 (10)	雪峰 (4, 2)、雲居 (4, 3)、保福 (1, 1)	長慶 (3)、三峰 (2)、阿彌陀佛 (2)、滄山 (2)、維摩 (2)
19	鏡清 (3-046-02)	62	雪峰 (23)、長慶 (6)、化度 (6)、		資福 (3)、釋迦 (2)、迦葉 (2)、招慶 (2)
20	夾山 (2-079-04)	58	韶山 (1)		佛日 (9)、洞山 (8)、欽山 (3)、鳳池 (2)、花亭 (2)、漳南 (2)、舍那 (1)

As can be seen, Chan masters like 仰山 Yangshan and 南泉 Nanquan have surprisingly high scores, while others like 馬祖 Mazu and 趙州 Zhaozhou seem surprisingly low, whereas 臨濟 Linji (37), 曹山 Caoshan (31) do not even enter the top 20. The latter however figures prominently in Nanquan's entry, while the former is suspiciously absent. Of the three founders of the Rinzai Sect and the forerunners of the Sōtō Sect, only 洞山 Dongshan (10) is listed here. This provides a good measure for the actual degree of influence the masters had on the development, as opposed to the image that was created of them in the early years of the Song<sup>24</sup>. It would be quite interesting to use this data and apply the methodology developed by Linton C. Freeman and others for social network analysis<sup>25</sup> to this data, but that is beyond the scope of this paper.

## 5. Conclusions

This paper tried to bridge the gap between texts of decidedly different genre, historiographical records of the Tang dynasty and a genealogical chronicle of Chan masters, which was seen as one example of how the content of the Knowledgebase could be extended beyond the basic records of the standard histories. It became apparent that the Knowledgebase as it stands now does not yet provide the means to meaningfully blend these texts into the Knowledgebase. Of the features that had been used in this comparison, the area of most overlap was in the reign names used for dates. Geographic names do also overlap to some degree, but more work needs to be done to allow for the less formal appellations used in the *Zutang ji*.

The most important area under investigation, that is the names of the persons that appear in these texts proved to be the area of least overlap: The difference in coverage of the standard histories and the Chan chronicle proved to be too big to be bridged easily. In this case, substantial work had been done to disambiguate and identify the persons that are mentioned in the texts. In order to bridge this gap, other means will be necessary, for example the inclusion of more diverse types of texts in the Knowledgebase, i.e. works of literature, epigraphic records, other administrative documents and so forth.

Whatever texts are added, however, the process of integration into the Knowledgebase will require a considerable amount of manual effort and the 'Knowledgebase System', the software for interacting with both the information layer and the resource layer will have to take this into account.

<sup>24</sup> This is quite a common place in recent Chan studies, (see Christian Wittern, *Style and Fashion in Early Song Chan Yulu*, in *Facets of Tibetan Religious Tradition*, edited by Alfredo Cadonna and Ester Bianchi, p141-142 and John McRae, *Seeing through Zen*, p76, 100 and 116-118.)

<sup>25</sup> See Thomas J. Fararo *Theoretical Sociology in the 20th Century*, in: *Journal of Social Structure* Vol. 2, 2 for an overview of the positions and Linton Freeman, *Visualizing Social Networks* in *Journal of Social Structure* Vol. 1, 1.. All articles of the *Journal of Social Structure* are online at <http://www.cmu.edu/joss/content/articles/volindex.html>

It has also become apparent during the research reported here, that the most universal framework applicable to all texts is the temporal framework of expressing dates. While presently the very rough approach of using the occurrences of reign names to identify dates has been used (and more detailed dates, or other ways of expressing a date - like "at the time emperor Xian Zong ascended the throne" have not even be marked yet), a more precise and more exhaustive markup of dates should provide a first framework that can be used to relate texts to each other and provide the necessary context to meaningfully disambiguate features of these texts. To a lesser degree, this holds also for geographic names, once the record of changes of these names has been integrated into the information layer. Together, these temporal and spatial features could provide the necessary context to relate the persons appearing in the text more accurately. Even the imprecise naming frequently seen like for example "Magistrate Lei 雷相公" could with the necessary temporal and spatial context quickly lead to the identification of this person.

In addition to this, more exhaustive listings of alternate names and orthographic variants should help to increase the recall rate. It would be useful to not only list exact matches to queries, but in the absent of such exact matches to give entries that at least partly match the desired criteria.

It is also quite obvious, that the simple notion of 'relation by co-occurrence' needs some refinement to be more widely usable. There are many reasons, why two items might occur together in a text and this precise relation is explicit and obvious in the text. To make this available to the kind of processing that is required for the Knowledgebase, the type of relation needs to be encoded more precisely.

Finally, a note on the technical side. The way the tables and figures in this presentations have been created is by way of using the 'XML Query Language'<sup>26</sup> to request the information and format them into a suitable way. This is a versatile and flexible way of processing this type of documents and it is planned to construct the interface to the Knowledgebase in a way that allows the users to make similar requests. The challenge will be to provide the powerful retrieval capabilities together with a intuitive and easily understandable user interface.

<sup>26</sup> A query language for XML (structured) documents, currently in the final stages of development at the World Wide Web Consortium. More information can be found at the website of the [XQuery workgroup](http://www.w3c.org/XML/Query) [<http://www.w3c.org/XML/Query>].

## **Historical Inscriptions at Cloud Dwelling Monastery: Preparations for a CD-ROM**

In more than five centuries the Cloud Dwelling Monastery (Yunjusi 雲居寺) at Fangshan produced about fifteen thousand stones on which texts from the Buddhist canon were engraved. To commemorate special events a number of historical inscriptions were written. Twenty-eight of these, dating from 616 to 1999, have been analyzed, transcribed, and translated.

To provide an easier access to the material, the texts shall be presented on a CD-ROM. This will pose a number of problems for both conception and design. Three of the most important are:

First, the texts have to appear in different forms, i.e. the user should be able to read the modern Chinese transcription parallel to the rubbing, and the annotated Chinese text parallel to the English translation. One will need to zoom into larger inscriptions or zoom out for a full view. Texts in the parallel windows optionally should be able to follow this movement or zoom.

Second, a lot of scholarship has already been done in reading the texts. As some of the stones only exist in fragments or have suffered by erosion, different readings and interpretations exist. The CD-ROM therefore has to provide the means to indicate these variants.

Third, the inscriptions reveal a lot of information on the huge project of cutting the sutras into stone. Many persons, for instance, are mentioned in no other known source. This makes it necessary for the CD-ROM to not only provide a full-text search, but to also offer special indices and background information.

The paper will introduce the user interface and discuss solutions for the three problem complexes mentioned above. It will also present the current state of the CD-ROM development and provide a first insight in this on-going project.

Matthias Arnold