

On the Awakening of the Buddhological Epigraphy and Philology from the AI

Haiyan Hu-von Hinüber^{1,2}, Sylvia Melzer^{3,4}

¹ Max-Weber-Kolleg, Steinplatz 2, 99085 Erfurt, Germany

² Shandong-Universität, Shanda-Nanlu 27, 250100 Jinan, China

³ Universität Hamburg, Centre for the Study of Manuscript Cultures, Warburgstr.26, 20354 Hamburg, Germany

⁴ University of Lübeck, Institute of Information Systems, Ratzeburger Allee 160, 23562 Lübeck, Germany

Abstract

This paper aims to define the requirements for the systematic study – with help of natural language processing and schema matching techniques - the Buddhist bronzes provided with inscriptions and scattered around the World. It concerns a pilot project dealing with 50-60 ancient Buddhist bronzes. Their inscriptions are written in Sanskrit language by using two different types of handwriting. According to the paleographic and historical studies, the scholarship has become able to assign these artifacts to the era of the royal family Paḷola Śāhi, which ruled the area Gilgit/Chilas and beyond, today in northern Pakistan, during the 6th – 8th centuries.

Keywords

Indology, information system, Buddhism, Post-Gandhāra, Sanskrit, Epigraphy, Paleography, Pakistan, Tibet, Beijing

1. Introduction

The royal family Paḷola Śāhi belonged to a dynasty of Buddhist kings in the Gilgit kingdom in the northern part of the Indian subcontinent in the 6th-8th centuries [1]. During this period roughly 50-60 ancient Buddhist bronze statues were manufactured with inscriptions written in Sanskrit language (see Figure 1). The Buddhist bronze statues were sponsored to someone to increase the quality of the person's rebirth. The Buddhism is an Indian religion or also philosophical tradition. Buddha means "awakened" and "is conferred on an individual who discovers the path to nirvana, the cessation of suffering, and propagates that discovery so that others may also achieve nirvana." [2] However, in the history of Buddhism, the goal of attaining nirvana was often considered unattainable in one lifetime. Therefore, the focus has been more on an accumulation of a good karma to increase the quality of rebirth. The increase in quality depends on the merits or demerits one has acquired through one's actions, as well as the merits a family member has acquired for one [3,4].

According to the paleographic and historical studies as published in last 20 years [5], the scholarship is able now to assign these special art objects of bronze foundry to the royal family Paḷola Śāhi of Iranian descent. In historical point of view, this area had been under the influence of Indian culture for several centuries. Therefore, it can be taken for sure that these bronzes originate from Historical Northwest India.

In order to make such statements as who sponsored which bronze statue to whom, how the statues were made, and what religious significance they had, data was often analyzed by hand in meticulous patient work and brought together from various fields of the humanities, as long as the connections

Humanities-Centred AI (CHAI), Workshop at the 45th German Conference on Artificial Intelligence, September 19-20, 2022, Trier, Germany
EMAIL: haiyan.hu.von.hinueber@orient.uni-freiburg.de (H. Hu-von Hinüber)

ORCID: 0000-0002-5284-9001 (H. Hu-von Hinüber)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

could be made. In addition, a major goal of Indologists is to find out the history of the Buddhist bronze statues themselves. However, this turns out to be very difficult because the available information is often incomplete.

As a current research question to be answered here is *why there are some bronze statues in Beijing whose origin can be attributed to the royal family Paḷola Śāhi*. For this purpose, it is necessary to evaluate the inscriptions of the above-mentioned Buddhist bronze statues e.g. from the field of epigraphy and philology. Epigraphy is the study of inscriptions to e.g. “clarify the meanings, classify their uses according to dates and cultural contexts, and drawing conclusions about the writing and writers.”[6] The Buddhist bronze statues have inscriptions, which are written in Sanskrit language by using two different types of the so-called “Gandhāra-Brāhmī” handwriting: the “round type” is from the 2nd century to the 630 AD, and the “rectangular type” form is from 630 to the 8th century. The study of the written language is called philology. The aim of this study is to determine the meaning of inscriptions. If the history in the 4th-6th centuries and the Buddhist bronze statue inscriptions are studied more closely, it is found that even the Tibetans consider the statues to be holy, but could no longer read and understand the writing. Therefore, in this paper we present the requirements to be solved by using AI methods to answer the research question by linking different research data sources.

	
<p>Front of the Bodhisattva Bronze from the Trashilünpo Monastery (Shigatse, Central Tibet) 12 cm high; pedestal measures 8,5 cm x 5,5 cm</p>	<p>Back with the Inventory Number of the Monastery bKra-“2588”</p>
	
<p>Left/Beginning of an inscription: <i>deyadharmo yam</i></p>	<p>Right/End of an inscription: <i>bandhuprabhāsasya</i></p>

Figure 1: A bronze statue with inscription kept in the Trashilünpo Monastery (Tibet) reading “This (statue) is a donation (given) by Bandhuprabhāsa”.

2. Representation and Retrieval of Research Data of Scholars from Different Disciplines and Countries

A special feature, or better to say, one of the special difficulties for the continuation of this project is due to the fact that only very few scholars, who have the special knowledge in old-Indian epigraphy, have worked on the Sanskrit inscriptions engraved in roughly 50-60 bronzes, namely on their reading, dating and assignment with help of the comparative analysis. Furthermore, the epigraphic and historical studies on the inscriptions should be supported by or cooperated with archaeologists excavating in North Pakistan and Tibet or often conducting field research. Thus, the relevant disciplines include Sanskrit Philology and Paleography, Archaeology (Pakistan and Tibet), Early History of Buddhism, as well as Buddhist Art and Epigraphy.

The researcher studying on the inscriptions and related bronzes work actually in different countries such as Germany, China, Switzerland, Italia, Japan and Holland (publishers). Nevertheless, these scholars from distant countries have to cooperate more closely, e.g. by linking via an information system:

- Oskar von Hinüber (Germany: indologist: one book and a dozen articles [5])
- Haiyan Hu-von Hinüber (Germany: indologist and editor of a new bronze information system)
- Luo, Wenhua & Team (China: archaeologists excavating in Tibet)

Some other cooperating scholars and institutions:

- Ulrich von Schroeder (Switzerland: pioneer; three volumes from 1987 and 2001)
- Kudo, Noriyuki (Japan: one of the main publishers)
- Jonathan Silk (Holland: one of the publishers)
- Elisa Iori & Luca Olivieri (Italia: archaeologists excavating in Pakistan)
- Different (private) art collectors and public museums

As far as the state of research is concerned, there are mainly four major projects/stages (2001, 2004, 2007–2018, and from 2022 on) to be considered [7-12]:

1. A number of inscriptions deciphered and translated into English by O. von Hinüber has been published by U. von SCHROEDER in his two volumes *Buddhist Sculptures in Tibet*, 2001. Most of the read inscriptions are also documented with photos.
2. The chapter “Inscriptions on Bronzes (no. 11–16)” and the “Addendum” in O. von HINÜBER’s monography *Die Palola Śāhis* (2004, pp. 28–42 & 190). The photos of all discussed inscriptions are published in this book.
3. From 2007 to 2018, O. von HINÜBER published a total of seven articles concerning inscribed bronzes originating from northwest India in the *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, vol. 10, 12–15, 18 and 21. The photos of the inscriptions are included in each respective volume.
4. In a close collaboration between Beijing and Freiburg which started in 2016, a number of inscribed Indo-Tibetan bronzes have been investigated within the framework of the paleographic and historical perspective. It concerns a dozen of newly found bronze sculptures that are equipped with Sanskrit inscriptions. As the “first-hand material”, these new finds examined in forthcoming publications (from 2022 on) can be regarded as supplements to the above-mentioned corpora documenting the Indo-Tibetan sculptures so far, especially the group of the so-called “portable statues” which do not always receive the greatest attention despite their large number. This German-Chinese collaboration forms a sub-project of the extensive survey project carried out by the Research Institute for Tibetan Buddhist Heritage of the Palace Museum (Beijing), which was set up ten years ago under the leadership of the institute’s director Luo Wenhua, one of the co-authors of this article.

During the cooperation with China in recent years (2016-2022), it turned out that the publications from Europe are either very expensive (von Schroeder 2001) or already out of print (von Hinüber 2004). It therefore makes sense to gather all relevant publications via the database, be it book chapters, monographs or individual articles.

Over the years, the researchers listed here have collected and analyzed data and stored the results in a wide variety of formats, i.e., printed books in pdf, XML, TIFF, DOCX, and CSV, or in databases with a non-standard, project-specific data model. From a technical point of view, the challenge now is to first store all data in such a form or to create an interface so that this data becomes exchangeable. From a humanities perspective, it must be determined which data can be mapped to each other or intelligent information retrieval (IR) systems deliver additional data which complement the data of the Bronze statues.

For the retrieval of additional data, we have already developed an algorithm, the Compl-IR algorithm [13], that returns additional IR results based on a similarity computation of data types of an entity. Entity types are e.g.: person, sponsor, organization, place, and date. The identification of named entities in a text and their classification into predefined categories (entity types) can be computed automatically using natural language processing (NLP) techniques such as named entity recognition (NER). The Open Information Extraction (OpenIE) annotator [14] can assign the entities, e.g. the family Paḷola Śāhi, mentioned in the texts to the entity types in which users are mainly interested in.

As a result, the Compl-IR algorithm can be used to obtain additional information to reconstruct the history of the royal family Paḷola Śāhi in more detail through the data of the Buddhist bronze statues, but also other artifacts.

Before the search algorithms can be used, the data must first become accessible. Thus, the first step is to build an information system that contains the data of the 50-60 bronze statues made in the 5-8 centuries, with the location Pakistan and inscriptions in the language Gandhāra-Brāhmī. An information system was already initially built with the database management tool Heurist [15] (see Figure 2). Heurist is an open-source database management system with a web-front end. Heurist allows researchers without prior IT knowledge to develop data models, store search, and publish data on a website.

The idea is to link this information system with the other existing information systems. For example, it is possible to link the database Indoskript 2.0 [7], which contains the letters from the South Asian region. However, only the letters used in Pakistan should be considered. An evaluation, which letters are relevant for the analysis of the inscriptions, should be computed automatically.

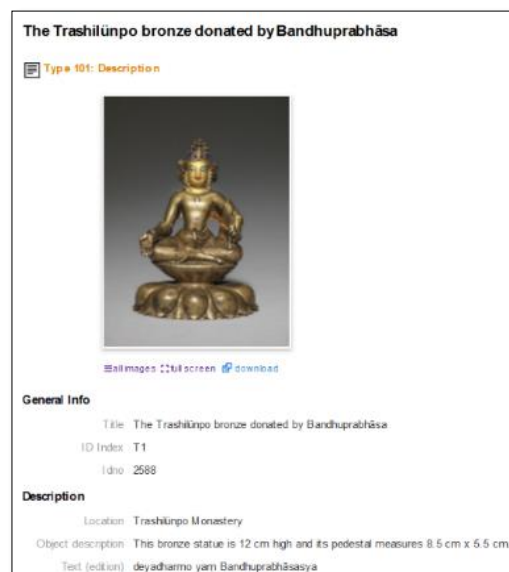


Figure 2: Information System built with the database management tool Heurist

3. An information system for paleographic analysis

Up until today, dating the Buddhist sculptures from historical Northwest India has always been a demanding challenge and in most cases even impossible, so that many questions concerning the date and the location of the origin as well as the tradition of the artistic style remain unanswered or have not been explained sufficiently. Therefore, the paleographic analysis is all the more significant to help in dating the bronzes - besides comparing the artistic style and iconography. Paleography is the study of ancient writing and inscriptions as well as the deciphering and interpreting historical manuscripts and writing systems [16]. It is concerned with the forms and processes of writing; not the textual content of documents. In the period from the 2nd to the 8th century, two main types of writing were used in Northwest India:

- 2nd century to 630 CE: the Gandhāra-Brāhmī or the so-called “round type”
- 630 CE to 8th century: the Proto-Śāradā or the so-called “rectangular type”

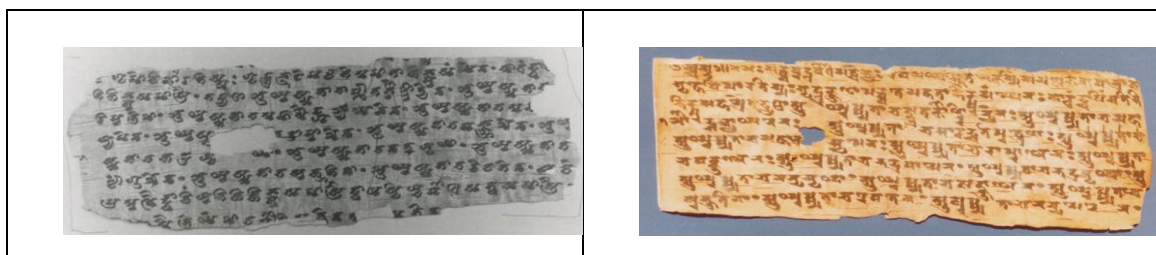


Figure 3: One folio from the Saṃghāṭasūtra found in Gilgit (Pakistan), written in the old type Gandhāra-Brāhmī (left); One folio from the Saṃghāṭasūtra found in Gilgit (Pakistan), written in the later type Proto-Śāradā (right)

In connection with setting-up a database for paleographic analysis, the following key points regarding the development of the script types as found in the Historical Northwest India should be considered:

1. The abrupt transition from the Gandhāra-Brāhmī “round type” (approx. 2nd–7th centuries) to the Proto-Śāradā type took place almost exactly around 630 CE.
2. This radical exchange of the script type seems to go hand in hand with the change in the title of the ruling family Paṭola Śāhis.
3. The script Proto-Śāradā gradually fell out of use, probably in connection with the decline of the Paṭola Śāhis family from around 740 onwards

Starting in 2005, Harry Falk (Prof. for Indology, FU Berlin) and Walter Slaje (Prof. for Indology, Uni Halle) established a paleographic database including different types of handwriting used in India from 3rd century BCE up today: Indoskript 2.0 [7].

However, there are **many gaps** that need to be filled, particularly in relation to the inscriptions found in Northwest India. Therefore, it would be useful to set up a database "Writings of Northwest India (2nd - 8th centuries)" which could benefit research in the longer term. Below are two examples:





			
Old type “ya” consisting of three parts	Later tepe „ya“ consisting of only two parts	Old type of the ligature “ndhu”	Later type of the ligature “ndha”

Figure 4: Two examples of handwriting occurred in Northwest India (6th -7th centuries)

4. Federated Indology Bronze Database System (FIndo BDBS)

The more databases exist, the more likely it is that additional data will be found to enrich the existing local information. To this end, federated searches and a federated database system (FDBS) provide users with additional information [17]. Users send their queries to the FDBS, which then forwards the queries to the individual (relevant) database nodes [18]. In order to be able to write down the history of the royal family Paḷola Śāhi in detail, the data of the bronze research are used as well as additional information is included. Additional information includes the writings used, the research data of wall paintings and manuscripts from the period when the royal family had ruled. A federated Indology bronze database system (FIndo BDBS) is created by federating databases which are in the field of Indology. In this paper, we want to highlight the requirements on how to systematically develop such a FIndo BDBS using AI methods.

- *Requirement 1:* The various research data sources need to be put into an analyzable form so that they can be shared with other research data.

Approach 1: Approaches such as transforming documents into a standard for databasing on demand can help ensure that data can be transferred to a database in a short time. In [19] it has been shown how to successfully transform a DOCX document to EpiDoc, a widely used data format in the field of Epigraphy, and then build a customized project-specific information system on demand.

- *Requirement 2:* Four types of heterogeneity (syntax, semantics, model, access) must be considered if the data schemas are to be matched and data of these schemas are to be mapped among each other.

Approach 2: When research data is stored in different databases, non-standardized, project-specific data models are often used. If a federated search is to be performed, it must be ensured that data schemas can also be matched and the data mapped to each other. There are a number of schema matching approaches that perform the process automatically, e.g. schema matching based only on the schema at structure- or element-level (linguistic, constraint-based), or based on the content at element-level (linguistic, constraint-based). [20] Deciding which paragraph would fit best depends on how the data is finally available.

We have already simulated schema matching of different data models [18] and determined that the precision must be very high in humanities' projects. Words can have a different syntax but the same semantics and vice versa. To the schema matching approaches we deal with topics such as name similarity, graph matching, and information retrieval. In [13] and [21], we have already show receive similar documents, an increase in precision can be achieved within a reasonable time frame, taking recall into account.

- *Requirement 3:* A technical infrastructure must be created that allows the various research data sources to be linked together.

Approach 3: In general, database federation provides logical centralization of data without the need to change physical database implementations. A common interface helps to create a basis for the exchange of data. In this case, it is a broker federation that allows the creation of messaging networks to which messages from one broker are automatically forwarded to another broker. RabbitMQ is an open-source message broker and provides broker federation between clients and servers. It has already been shown that cross-domain information systems can be generated using RabbitMQ and federated search queries can also be executed. [22]

If these requirements are implemented, a FIndo BDBS is realized. Some of the approaches were tested individually and with other data sets from different projects. For the new project, new data sets (Gilgit Buddhist Inscribed Bronzes) will be used to obtain new information on the history of the royal family Paḷola Śāhi via AI methods.

5. Conclusion

This paper gives an overview about the requirements for the systematically study of the Buddhist bronzes, which are provided with a Sanskrit inscription and originate from the historical Northwest India (today Pakistan). An information system was built with the tool Heurist to describe and represent the history of the royal family Paḷola Śāhi and their culture. With the linking to other sources of information a FIndo BDBS can be realized to come to new knowledge. For the realization of a FIndo BDBS it belongs to deal with different schema matching techniques and to use them in this new overall context. The approaches already implemented show that this approach can be promising for the new project.

6. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796

7. References

- [1] von Hinüber, Oskar, "Bronzes of the Ancient Buddhist Kingdom of Gilgit", University of Freiburg, <https://www.metmuseum.org/metmedia/video/collections/asian/bronzes-of-ancient-gilgit>
- [2] Siderits, Mark, "Buddha", The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, 2019.
- [3] Buswell, Robert E. Jr., Lopez, Donald Jr. (2003), "The Princeton Dictionary of Buddhism", Princeton University Press, 2003
- [4] Ronald Wesley Neufeldt, "Karma and Rebirth: Post Classical Developments", State University of New York Press, pp. 123–131, 1986, ISBN 978-0-87395-990-2.
- [5] Von Hinüber, Oskar, "Publikationen", <http://www.iriaabs-freiburg.de/index.php/2-uncategorised/8-oskar-von-hinueber#publikationen>
- [6] "Epigraphy", 2022, <https://en.wikipedia.org/wiki/Epigraphy>
- [7] Falk, Harry, Slaje, Walter "Eine elektronische indische Paläographie (Programmierung: O. Hellwig; Dateneingabe: K. Einicke, K. Hoffmann, J. Neuß), Berlin 2005. <http://userpage.fu-berlin.de/falk>.
- [8] Wang, Xin, Tapani Ahonen, and Jari Nurmi. "Applying CDMA technique to network-on-chip", IEEE transactions on very large scale integration (VLSI) systems 15,10, 2007, pp. 1091-1100.
- [9] P. S. Abril, R. Plant, "The patent holder's dilemma: Buy, sell, or troll?", Communications of the ACM 50, 2007, 36–44. doi:10.1145/1188913.1188915.
- [10] Hellwig, Oliver, "Dating Sanskrit texts using linguistic features and neural networks", in: Indogermanische Forschungen 2019, <https://www.degruyter.com/document/doi/10.1515/if-2019-0001/html?lang=de>
- [11] von Hinüber, Oskar, Luo, Wenhua, "The inscribed Buddha image donated by Vappaṭa and Dhruvabhaṭā kept in the Sakya Monastery", in: Annual Report of The International Research Institute for Advanced Buddhology at Soka University for the Academic Year 2021, Vol. 25, Tokyo 2022: 3-9, plates 1 - 4 (8 figures).
- [12] Hu-von Hinüber, Haiyan, Luo, Wenhua, "Two Newly Found Bronze Statues with Sanskrit Inscriptions originating from Historical Northwest India", Connectiong the Art, Literature, and Religion in South and Central Asia. Studies in Honour of Monika Zin, ed. by I. Konczak-Nagel, S. Hiyama and A. Klein, Delhi 2022: 161-170 (with 10 figures).
- [13] S.Melzer, S. Schiff, R. Möller, "Complementary Document Representations for Information Retrieval", The 34th International FLAIRS Conference, North Miami Beach, Florida, USA, 2021
- [14] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, "The Stanford CoreNLP natural language processing toolkit. Proceedings of 52nd Annual Meeting of the ACL: System Demonstrations, 55–60, Association for Computational Linguistics, 2014

- [15] Heurist (2022) “A unique solution to the data management needs of Humanities researchers.” Software available: <https://heuristnetwork.org/>
- [16] ”Palaeography, n.”, Oxford English Dictionary (Online ed.). Oxford University Press. (Subscription or participating institution membership required.)
- [17] Melzer, Sylvia. (2022, March). Federated Search in Manuscript Databases. <http://doi.org/10.25592/uhhfdm.10289>
- [18] Melzer, Sylvia, Thiemann, Stefan, Möller, Ralf ”Modeling and Simulating Federated Databases for early Validation of Federated Searches using the Broker-based SysML Toolbox”, The 15th Annual IEEE International Systems Conference (SYSCON 2021), virtual conference, 2021
- [19] Melzer, S., Schiff, S., Weise F., Harter, K., Möller, R. ”Databasing on demand for research data repositories explained with a large epidoc dataset”, CENTERIS 2022 - Conference on ENTERprise Information Systems, 2022 (will be published)
- [20] Rahm, E., Bernstein, P. A survey of approaches to automatic schema matching. The VLDB Journal 10, 334–350 (2001). <https://doi.org/10.1007/s007780100057>
- [21] S. Melzer; Semantic Assets: Latent Structures for Knowledge Management, University of Lübeck, 2018, phd thesis
- [22] S. Melzer, H. Peukert, H. Wang, S. Thiemann, ”Model-based Development of a Federated Database Infrastructure to support the Usability of Cross-Domain Information Systems”, The 16th Annual IEEE International Systems Conference (SYSCON 2022), Montreal, Canada, 2022