Original Research Paper

Historical Relation Extraction from Buddhist Temple Documents of the Lanna Kingdom

[1]Pongkorn Chantaraj, [1]Jaratsri Rungrattanaubol and [2]Anamai Na-udom

[1]*Department of Computer Science and Information Technology, Faculty of Science,*
Naresuan University, Phitsanulok, Thailand
[2]*Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok*

Abstract: A substantial amount of valuable knowledge has been recorded, namely historical documents, in the form of unstructured data. The Extraction of such information will provide useful historical knowledge. The objective of this study, is to propose a relation extraction technique of unstructured data from Buddhist temple documents in the Lanna Kingdom, consisting of relationships including the name of the kings who built the temple, the name of the king who created or restored a historical site within its grounds and the Buddhist era the temple was built. The research process began with collecting temple data located in the upper northern part of Thailand, with published documentation of temple history in electronic form from Temples Thailand online website. The data was transformed into an electronic document or a text file, then passed to the LEXTO program for word segmentation, while retaining the original meaning. We designed and developed an algorithm to extract relationships between named entities from Buddhist temple documents, and the application of information extraction principles to the process. We then evaluated the 1,093 original documents by using the algorithm to extract the required information. From the research, the results indicated that the proposed technique can be applied effectively, to accurately extract the relations related to Lanna kingdom history from Buddhist temple documents.

Keywords: Information Extraction, Relation Extraction, Named Entity Recognition, Historical Relation Extraction, History of Lanna Kingdom

Introduction

The Lanna Kingdom is one of the most important regions of Siam, known as Thailand today. Lanna was established in 1296 by Phraya Mangrai, the 1st king of Lanna, who founded Chiang Mai as the capital city and the center of government and culture. Chiang Mai, at present is one of the most attractive cities in the Northern part of Thailand, and is the major center of Lanna culture and civilization. Many historical remains can be found in Chiang Mai and the cities nearby, such as ancient city walls, pagodas, Buddha paintings and statues, and hundreds of Lanna temples that have stood for half a millennium or more.

The Lanna Kingdom became part of Thailand in 1396 (Penth, 2001), and was proclaimed as "The Land of Buddhism". It is one of the best destinations for people who want a better insight into the teachings of Buddhism and the chance to experience life in a Buddhist monastery.

During the Sukhothai Kingdom period, Buddhism was accepted as the national religion by the king, and has been closely related to Thai people since 1257. Since then, most of the Thai population have become Buddhists and followed the teachings of Buddha (Piyabhani, 2016).

The Lanna Kingdom was a loyal supporter of the Sukhothai Kingdom since the early period, so Buddhism flourished and influenced the Lanna society in many aspects, making them practically inseparable (Wonglangka and Han, 2018). A Buddhist temple (called "wat" in Thai) is a place to perform sacred rituals and practices, and also a place for education and vocational training.

From past to present, historical sites such as Buddhist temples and religious buildings have been designated as the center of a community and a place to display valuable artifacts (e.g. Emerald Buddha statue, Jed Yod pagoda etc.), which were maintained by kings, noble citizens, rich merchants, and the general public.

Since the early age of the Lanna Kingdom, the kings had to be ordained as monks to uphold Buddhism, while establishing and maintaining religious buildings. The information written in Buddhist temple documents are partially related to the history in some way, however, most temple documents come in various forms, such as inscriptions, myths, and memorandums including archaeological evidence of pagodas, chapels, and Buddha images. They were written in the old-style Thai language, and may contain some drawings. Most Thai historians study the various forms of documents and related materials in order to extract trustworthy history and make them academically sound under the authorship or editorship of recognized experts. This process is time-consuming which requires individual knowledge and expertise.

The Buddhist temple documents, the remarkable artifacts and remains in a temple can be studied, interpreted and extracted with respect to Lanna history. The Lanna kingdom is located in the upper northern part of Thailand and used to control some border lands of Laos and Burma. There are a thousand Buddhist temples and historical sites located around this area, and the study of these can significantly reveal both the past and present history of Lanna.

Historical evidence of temples in Thailand are collected by the National Office of Buddhism, the Fine Arts Department and Temples Information Center. The history of temples and related artifacts have been studied and written about by experts, then stored in book form, while some are transformed into a digital format for websites. However, most of the documents are in text, which are unstructured with no classification. The information required from this data can only be gleaned from the original documents, which is time consuming and causes delays in the search. To counter this problem, if the documents could be processed using the information extraction technique and systematically store the information, this would create a knowledge base that would be useful in a historical sense, and allow people in society to recognize and understand the contents.

Based on the historical concept, the name of kings, remarkable historical events, and valuable artifacts can denote the historical point of view. Hence, this research focuses on extracting the historical relationships between the Lanna kings and the Buddhist temples, such as the creation of a temple or the creation/restoration of a historical site within its grounds. This would show stability, prosperity and peaceful coexistence within the society, with the monarchy being the center of the public's faith from past to present.

At present, most historical works and data are from the knowledge and skills of experts who study and analyze related materials, and finally produce the reports. For instance, Baker and Phongpaichit (2014) published the history of Thailand from the early settlements in the Chao Phraya basin till now. Bovornkitti (2005) studied the historical relationship between Buddhism and indigenous Thai art, which led to the influence of Buddhism on the development of the current Thai art, and Kirigaya (2014) presented the important events of the Lanna era under the rule of Myanmar, through Burmese documents and Lanna chronicles. This paper foresees the possibility to extract some historical aspects from Buddhist temple documents automatically based on information extraction techniques, instead of solely relying on experts.

Information extraction is a process used to extract name entities and the relationships between them from articles, normally written in natural language. Its technique is based on text mining, using word analysis from unstructured data by specifying references to named entities and the relationships between them, and ultimately obtain the knowledge in return. It has been applied in diverse areas such as Hao *et al.* (2017) developed an automated verb-based algorithm for multiple-relation extraction on biomedical entities from PubMed database, Abraham *et al.* (2018) proposed the spatio-temporal extraction for historical events from text documents of the German–Herero resistance war to help efficient historical knowledge management, and Cybulska and Vossen (2011) introduced historical event extraction in Dutch historical documents from Srebrenica corpus to represent event actions with their participants, locations and time markers.

This research proposed the use of the information extraction technique on Buddhist temple documents which are unstructured, in order to extract historical relationships regarding the Lanna Kingdom, and the names of Lanna kings were used as the historical timeline. There are, in fact, many relationships that the kings were involved in, which are mentioned in the temple documents. However, in this paper we aimed to perform three historical relation extractions; the Lanna king who built the temple, who created or restored a historical site within its grounds, and the Buddhist era of the temples creation. The extracted relation results are further used in developing the Lanna history web-based application that will support a chronological search on the history of the Lanna kingdom, therefore, a high accuracy of extraction is required.

The remainder of this paper is organized as follows. First, we discuss some related works in the field of information extraction and historical knowledge, then, describe the process to extract the relationship between named entities from Buddhist temple documents, and

lastly, discuss and report the experimental results, and conclude by summarizing the key points of our work, and presenting future directions.

Related Works

A process of information extraction is commonly used to refine structured data or knowledge from unstructured text, by identifying references to named entities and stating the relationships between them. It has been applied to extract knowledge from the corpus or to extract data from documents accurately (Mooney and Bunescu, 2005). Apart from the English language, information extraction has been extended to extract data in other languages, which have different techniques to serve their natural languages.

Sazali *et al.* (2016) proposed a method of extracting nouns with named entity recognition from classical Malay documents, using the morphological rules (verb, adjective and noun affixes), in which the results were verified by language experts. A similar research by Sharum *et al.* (2011) who presented a way to recognize names in the Malay text of newspaper articles, which were published on the Utusan website. It was found that a "name indexer" application can be used to search for person's names in newspaper articles, but it is not capable of covering all names in Malay text without further comprehensive rules or patterns.

Zayed and EI-Beltagy (2012) proposed extracting the names of males and females from modern standard Arabic colloquial text, to create a dictionary of Arabic family names. Aboaoga and Ab Aziz (2013) also presented a rule-based approach to recognize person's names from Arabic corpus based on newspaper archives, while, Saif and Aziz (2011) used the hybrid method, based on linguistic information and association measures, to extract Arabic collocations from online newspaper archives.

Nakatoh *et al.* (2011) proposed relation extractions of place names from tourism blogs related to Kyushu area in Japan. This was carried out to make clarifications of places of interest and sightseeing locations clearer for tourists.

Chen and Hsu (2016) proposed multiple instance learning to extract concept pairs from the Chinese corpus, resulting in the reduction of human resources for processing the information.

In Thai, there are many proposed techniques for extracting information and relations such as Aunanan *et al.* (2016) proposed extracting Thai news contents by auto-tagging for faster information retrieval, and Sutheebanjard and Sutheebanjard (2009) proposed the method to extract Thai personal named entities without using Word Segmentation or POS Tagging from news articles which greatly reduces the time and effort used in building the training corpus, while, Tongtep and Theeramunkong (2009) presented a feature-based approach for extracting relations between a pair of

named entities (person name, location, and action) from Thai news documents, which gave accurate results. The major problems of information extraction from Thai text is caused by long sentences with no use of punctuation and the lack of use of upper and lowercase letters to identify named entities, and the inconsistency of the writing (Tepdang *et al.*, 2010; Wikaha and Netisopakul, 2014).

In the case of history, researchers focus on gathering historical data and compiling reports, presentation and visualization of their results. For instance, Biller *et al.* (2013) developed a web-based system for ground truth generation, to help historians and computer scientists collectively record annotate historical documents, while, Hasan *et al.* (2013) proposed a model to explain Historical events with temporal data using a timeline visualization, and Kauppinen *et al.* (2008) applied ontology time series to create an application for information retrieval and visualization in a semantic cultural heritage portal.

As mentioned before, most Thai historical extractions are based on human experts and in a form of reports of particular historical interests (Bovornkitti, 2005; Baker and Phongpaichit, 2014; Kirigaya, 2014). Researches have hardly ever discussed how to extract information automatically to obtain historical facts.

Hence, we proposed a historical relation extraction method to obtain previously unobtainable knowledge from temple documents, which can be used in the web-based application supporting the history of the Lanna Kingdom.

Materials and Methods

In this paper, we proposed the extraction of three historical relationships from Buddhist temple documents. By using the Lanna Kings to denote the historical viewpoints and Buddhist era or calendar, three relationships of interest here are "The name of the king who built a temple" labelled as R1, "The name of the king who created or restored historical sites within its grounds" as R2, and "The Buddhist era the temple was built" as R3, as shown in Fig. 1. The names of Lanna Kings, the artifacts in the historical sites such as pagodas, tunnels, and chapels are the main collection of named entities, and a collection of keywords was designed for extracting the relations.

Based on the information extraction principles along with the proposed algorithm, the entire process for historical relation extraction is displayed in Fig. 1. The process is divided into two main steps, which are data collection and text processing for information extraction. The following sections describes each of these steps in detail.

Data Collection

In this study, 1,093 Buddhist temple documents were used as a dataset and all of them are Thai text-based. The temples were built during the Lanna period from 1261 to 1939 (Ongsakul, 2005) and are located in the upper northern part of Thailand.
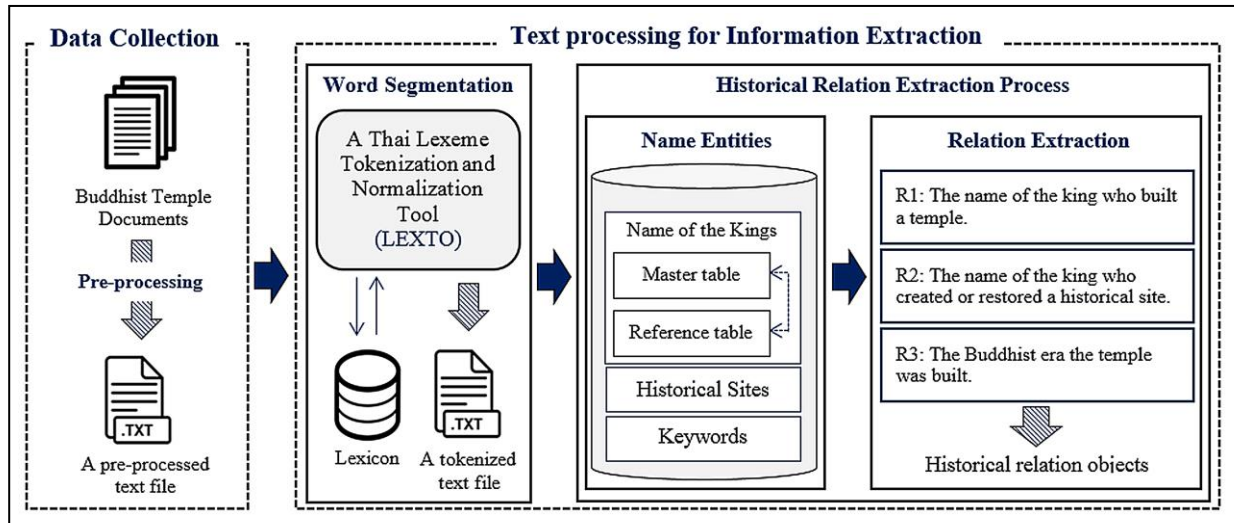
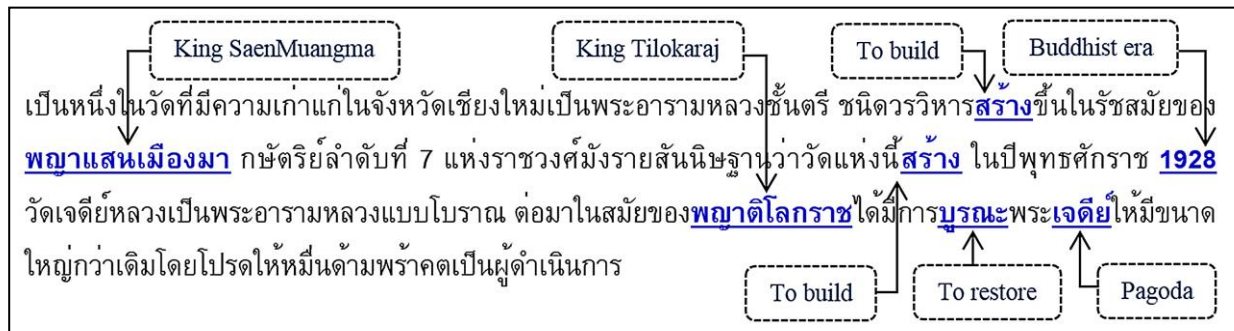**Fig. 1:** The proposed process to extract the relationships from Buddhist temple documents

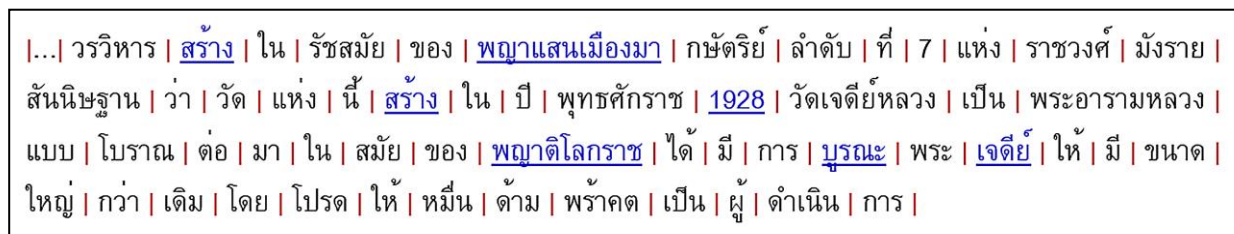**Fig. 2:** An example of the temple documents

Fig. 3: A tokenized file resulted from LEXTO

The documents are from the Temples Thailand online website (TIC, 2016). A sample of a document is shown in Fig. 2. The underlined words are examples of named entities and keywords that are fundamentally useful for historical relation extraction. The relations that may be extracted from the temple document in Fig. 2, are shown as follows:

<The temple> <built by> <*King Saenmuangma*>
<*King Tilokaraj*><restore><Pagoda>
<The temple> <built in> <1928>

In regards to the Thai language having no delimiters, spaces or full stops, the process requires effective word segmentation.

Text Processing for Information Extraction

The main objective of this step is to identify the relationships between named entities in Buddhist temple documents. After preprocessing all documents by removing unwanted or special characters, they are then passed through the word segmentation and relation extraction process.

Word Segmentation

Word Segmentation in Thai is a very difficult task since there are no distinctive clues to help such as spacing and punctuation, as in English, therefore the Thai language requires effective tools for accurate data processing. In this research, we use a Thai Lexeme Tokenization and Normalization Tool (LEXTO) developed by the National Electronics and Computer Technology (Haruechaiyasak and Kongthon, 2013). The preprocessed documents are entered into LEXTO to separate or tokenized each word in the file, but retain the original meaning to be used in the relation extraction process. The result, a tokenized file, is shown in Fig. 3.

Historical Relation Extraction Process

This section explains how to design an algorithm to extract historical relationships between designed named entities from Buddhist temple documents. The Lanna king is selected to represent a historical timeline in this study, since each king has his own reign period. There are 42 king's names in the Lanna kingdom (Ongsakul, 2005).

The artefacts or historical sites inside the temples are marked as another set of name entities, in which the king may be involved.

A set of keywords, which mainly are verbal, indicate and identify references between the relationships of the named entities.

The historical relation extraction based on the name of a Lanna king and Buddhist era, is divided into 3 objectives of extraction:

R1: The name of the king who built a temple,
R2: The name of the king who created or restored a historical site, and

R3: The Buddhist era the temple was built.

From preliminary study of Buddhist temple documents, it shows that more than one names could refer to the same king i.e. alias name. For instance, four different spellings are referred to King Mangrai as shown in Fig. 4. Therefore, we designed the table for the name of a king into two tables (Master and Reference Table). The total number of Lanna kings is 42, and the added aliases are 102.

As mentioned before, the relations extracted here are involved with some buildings in the temple grounds. Hence, the designed artefacts in the historical sites such as pagodas, chapels, main halls, scripture halls, walls, pavilions etc. are a set of named entities that the Lanna king may have been involved in creating or restoring them as shown in Table 1. The total number of the historical sites is 23.

Finally, as designed, the set of keywords representing the action are shown in Table 2. The total number of keywords is 15, which are divided into two groups to determine the meaning of "to build" or "to restore".

Table 1: Example of historical sites

Historical sites	Meaning
เจดีย์	Pagoda
วิหาร	Chapel
อุโบสถ	Main hall
หอไตร	Scripture hall
กำแพง	Wall
ศาลา	Pavilion

Table 2: Example of keywords

Keywords	Meaning	Grouping
สร้าง	to create / build	to build
สถาปนา	to establish	to build
ก่อตั้ง	to founded	to build
บูรณะ	to restore	to restore
ซ่อมแซม	to recondition	to restore
เสริมสร้าง	to renovate	to restore
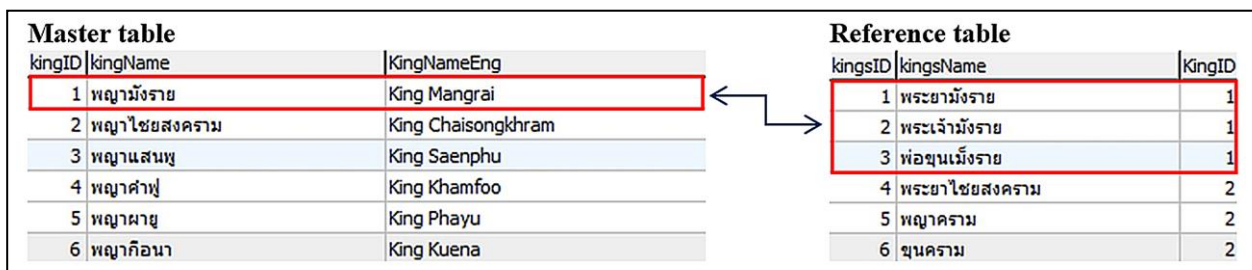
Fig. 4: The tables for the names of kings

The historical relation extraction algorithm is divided into 3; R1, R2, and R3 as mentioned earlier. Two relationships, i.e. R1 and R2, are related to the name of a Lanna king, while R3 is the period of the temple creation. The core proposed algorithm is written as a pseudo code displayed in Algorithm 1.

Algorithm 1: The core algorithm for historical relation extractions

---

1. **Begin**
2. **Set** arrayKings ← Read data from *Name of the king*
3. **Set** arrayHistoricalSites ← Read data from *Historical sites*
4. **Set** arrayKeywords ← Read data from *Keywords*
5. **Set** arrayHistory ← Read tokenizedTextfile.txt – *Temple document*
6. **Set** templeName ← Read a FileName
7. **for** i=0 to length(arrayKings)
8.   **for** j=0 to length(arrayHistory)
9.   **if** arrayHistory[j]==arrayKings[i]
10.   *R1: searchBuiltByKing(j)*
11.   *R2: searchCreateRestore(j)*
12.   **else**
13.   "no relations R1 and R2 found"
14.   **for** k=0 to length(arrayHistory)
15.   **if** arrayHistory[k] is 4 digit or 2 digit
16.   **if** arrayHistory[k-1] is "พุทธศักราช" or "พุทธศตวรรษ"
17.   *R3: searchBuiltInEra(k)*
18.   **else**
19.   "no relation R3 found"
20.   **End "algorithm"**

The core algorithm begins by reading the names of kings, historical sites and keywords and places them into a form of array; *arrayKings*, and *arrayHistoricalSites*, *arrayKeywords*. After that, the tokenized text file resulted from the word segmentation step is read into *arrayHistory*. Start by searching for the name of the king in the *arrayHistory*, then if the name is found, pass the position of it to the method process *searchBuiltByKing* and *searchCreateRestore*, otherwise no relations for R1 and R2 are in this document. The algorithm will search for the date and a keyword that will denote the Buddhist era the temple was built. The details of each relation extraction are discussed in the following sections.

Relation Extraction of the name of the King who Built a Temple (R1)

The name of the king who build the temple and the history aspect of it, can be used in the study of Lanna history. The relation extraction R1 is written in the following form:

<TempleName>, <Keyword>, <KingName>

By studying the temple documents, we found that the position of the keyword "to build" is normally located within the first 10 words, either before or after the name of the king, as depicted in Fig. 5.

The designed method for R1 is to search the set of keywords in the "to build" group, in *arrayHistory* within a range of 10 words, either before or after the position of the name of the king, and if found, we need to recheck whether there is a historical site mentioned. If one is found, it means that "to build" is not meant to describe the temple, but rather the historical site as shown in a dotted rectangle in Method for R1. Where pos represents the position of the kings name, it is then passed from the core algorithm to the *searchBuiltByKing(pos)*.
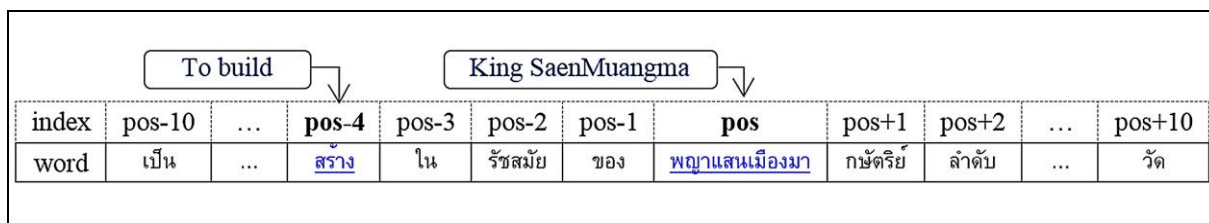
Fig. 5: A position of the name of the king and keyword in the document
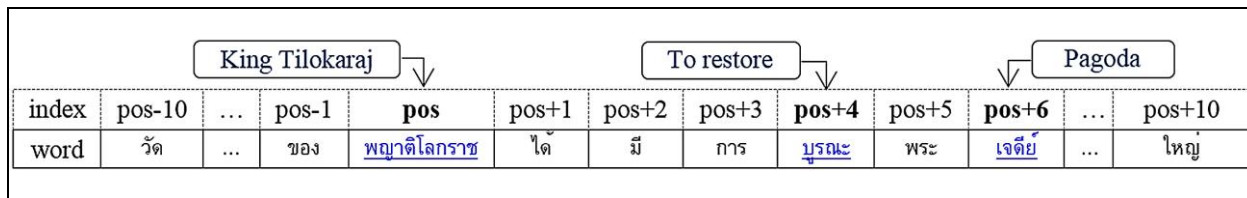
Fig. 6: The position of the name of the king, keyword and historical site that appears in the document

1325

index	pos-10	...	**pos-4**	pos-3	pos-2	pos-1	**pos**	pos+1	pos+2	...	pos+10
word	มังราย	...	สร้าง	ใน	ปี	พุทธศักราช	1928	วัดเจดีย์หลวง	เป็น	...	ของ

To build → pos-4   Buddhist era → pos

Fig. 7: The position of year and keyword in a document

Method for R1: searchBuiltByKing(*pos*)

for m=pos-10 to pos+10
 for n=0 to length(arrayKeywords)
  if arrayHistory[m]==arrayKeywords[n]
   for p=pos-10 to pos+10
   for q=0 to length(arrayHistoricalSites)
    if arrayHistory[p]==arrayHistoricalSites[q]
     "not for the temple, but for historical sites"
    write templeName, arrayKeywords[n],
       arrayHistory[pos]
  else
   "no relation R1 found"
End "Method for R1"

The Relation Extraction of the name of the King to the Creation or Restoration of a Historical Site Within the Temple Grounds (R2)

In this section, we wanted to extract the relationship of the king, who created or restored any historical site within the temple grounds who was mentioned in the document. The relation extraction is written in the form:
<TempleName>,
<HistoricalSite>,<Keyword>,<KingName>

The method of R2 is similar to R1, since it still concentrates on the name of the king. R1 aims for the king who built the temple, whereas R2 focuses on any mentioned king involved in creating or restoring any historical sites within temple grounds as shown in Fig. 6.

The method for R2 is almost the same as R1, except for the keywords covering both groups "to build and to restore", and the wanted relations for R2 are depicted in the dotted rectangular area.

Method for R2: searchCreateRestore(*pos*)

for m=pos-10 to pos+10
 for n=0 to length(arrayKeywords)
  if arrayHistory[m]==arrayKeyords[n]
   for p=pos-10 to pos+10
   for q=0 to length(arrayHistoricalSites)
    if arrayHistory[p]==arrayHistoricalSites[q]
     write templeName, arrayHistoricalSite[q],
        arrayKeywords[n], arrayHistory[pos]
   else
    "no relation R2 found"
End "Method for R2"

The Relation Extraction of the Buddhist Era the Temple was Built (R3)

There are two types of documentations for the year temples were built, which consists of the Buddhist Era (พุทธศักราช) which is documented by a 4 digit number, such as "พุทธศักราชที่ 1839", and the Buddhist Century (พุทธศตวรรษ) only has a 2 digit number, such as "พุทธศตวรรษที่ 22". The Buddhist Century is converted into the Buddhist Era with a straightforward formula e.g. Buddhist Century 21 is Buddhist Era 2001 – 2100, 22 is 2101-2200, etc.

Instead of finding the position of the name of the king, the core algorithm detects the position of the numbers and passes this on to the *searchBuiltInEra(pos)* as shown in Fig. 7. If the word proceeding the numbers are not "Buddhist Century" or "Buddhist Era", there is no R3, as shown in line 15-line 19 in Algorithm 1.

Similar to the previous methods, R3 starts by searching for the keyword "to build" within a 10 word range before or after the position of the numbers indicating the Buddhist era. If the keyword "to build" is found in *arrayHistory*, we need to recheck that the number is not for a historical site. If the year timeline of the discovery is in the "Buddhist Century", the system will automatically convert the numbers into the "Buddhist era". For example, the Buddhist Century 22 would be converted to the Buddhist era, between 2101 and 2200. The result are displayed in the following form:
<TempleName>, <Keyword>, <Year>

The extraction R3 represents the year of temple construction.

Method for R3: searchBuiltInEra(*pos*)

Set year <- arrayHistory[pos]
for m=pos-10 to pos+10
 for n=0 to length(arrayKeywords)
  if arrayHistory[m]==arrayKeywords[n]
   for p=pos-10 to pos+10
   for q=0 to length(arrayHistoricalSites)
    if arrayHistory[p]==arrayHistoricalSites[q]
     "not for the temple, but for historical sites"
   if year is a 2 digit number
     convert(year)
   write templeName, arrayKeywords[n], year
  else
   "no relation R3 found"
End "Method for R3"

Fig. 8: Examples of the results for the relation extraction of R1, R2 and R3

Figure 8 displays examples of historical relation extractions, which are formatted in the designed form as explained previously. Each temple document has at least three lines for relational extractions for R1, R2, and R3. There are three temples in the example, in which temple1 has 5 lines of extractions (line 1 – 5), and temple2 and 3 have 3 lines, which can be interpreted based on each relational extraction. R1, the king who built a temple is extracted as shown in line 1, line 6 and line 9, where **null** in line 9 indicates that no king who built the temple is identified.

R2, the king created or restored a historical site is listed in line 2 to 4, line 7 and line 10, where **null** in line 10 indicates that an R2 relation is not found. Line 2 to 4 illustrates there is more than one R2 relation in the temple document.

If the period of time the temples were created exists in the extractions for R3, the results will be displayed in lines 5, 8 and 11. If the year is a 4 digit number, i.e. Buddhist era, it will be displayed in line 5, but if the number only has 2 digit then this will also be converted into the Buddhist era, as shown in line 11, while **null** in line 8 notifies that the year of the temple creation is not found in the document.

The proposed algorithm is validated and measured for its accuracy by comparing the extracted results with the actual answers that appeared in the original documents. The next topic will explain how to set up the experiment and then measure the accuracy and performance of the algorithm.

Experiments and Evaluation

The experiment is set up to evaluate the proposed algorithm for the three designed relationships relating to Lanna Kings from the Buddhist temple documents and the Buddhist era of the temple construction. The three relation extractions are stated as R1, R2 and R3, as shown in Fig. 1.

This research used a combination of an information extraction technique and existing tools; Thai Lexeme Tokenization and Normalization Tool (LEXTO) used for Thai word segmentation. The proposed algorithm is implemented in Java programming language, and the Buddhist temple dataset is 1,093 documents from the Temples Thailand online website. The evaluation method was done by comparing the results extracted from the system with the actual answer appearing in the original documents. The performance of the experiment is measured with Precision, Recall and F-measure criterion:

- *Precision* is the percentage of the results which are relevant to the actual answer.
- *Recall* refers to the percentage of total relevant results correctly classified by the proposed algorithm.
- *F-measure* is the percentage of the harmonic mean of precision and recall.

They are formulated as Equation 1 to 3.

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} \times 100 \tag{1}$$

$$\mathrm{Re}call = \frac{TP}{TP+FP} \times 100 \tag{2}$$

$$F-measure = \frac{2 \times \mathrm{Pr}ecision \times \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} \times 100 \tag{3}$$

In this research, the meaning of the TP, FP, TN and FN are defined as follows:

True Positives (TP): refer to a complete list of named entities in the documents and accurate extraction by the system.

False Negatives (FN): refer to a list of named entities in the documents, but the system give inaccurate result or **null**.

True Negatives (TN): refer to no list of named in the documents entities, and accurate extraction as a result of **null**.

False Positives (FP): refer to no list of named entities in the documents, but the system give inaccurate result as a name entity.

The performances of the algorithms are shown in Table 3.

Table 3 shows the performance of the algorithms in terms of precision, recall, and F-measure. The algorithm for extracting R1, performed best with the highest score of all three criteria at over 99% each, while R2 performed the second best, followed by R3.

From the study, there were some incorrect extractions in R1 and R2, because some names that were meant to be Lanna kings in the documents had not been included in the list of names, and some keywords meaning temple construction in the documents, were also not in the system. Although these problems can be resolved simply by adding other alias king names and keywords into the system, however, to completely cover all related names and keywords in the system is quite hard to implement, therefore, we just keep adding names and keywords when they are encountered.

As for R3, it had the lowest rate of 94.14% on F-measure, because some documents contained more than one date in the vicinity of the keyword, while some of the dates showed no relation to temple construction, but did show a relation to antiques within the temple instead, which caused an error in the relation extraction.

Some other problems found during the experiment were some misspelled words, some complex sentences, and multiple relationships between named entities. This caused the extraction process to have errors which made handling these kinds of documents in the Thai, difficult.

However, our process can extract all relations with an overall accuracy of over 94%. The historical relation extractions from this study can be further used in developing an application that supports Lanna history as depicted in Fig. 9. This application provides chronological search related to the Lanna kingdom with a connection between the location and the time period based on the Lanna King era.

Table 3: The relation extraction performance

Relation	Precision (%)	Recall (%)	F-measure (%)
R1	99.10	99.06	99.08
R2	98.69	98.44	98.56
R3	94.17	94.11	94.14
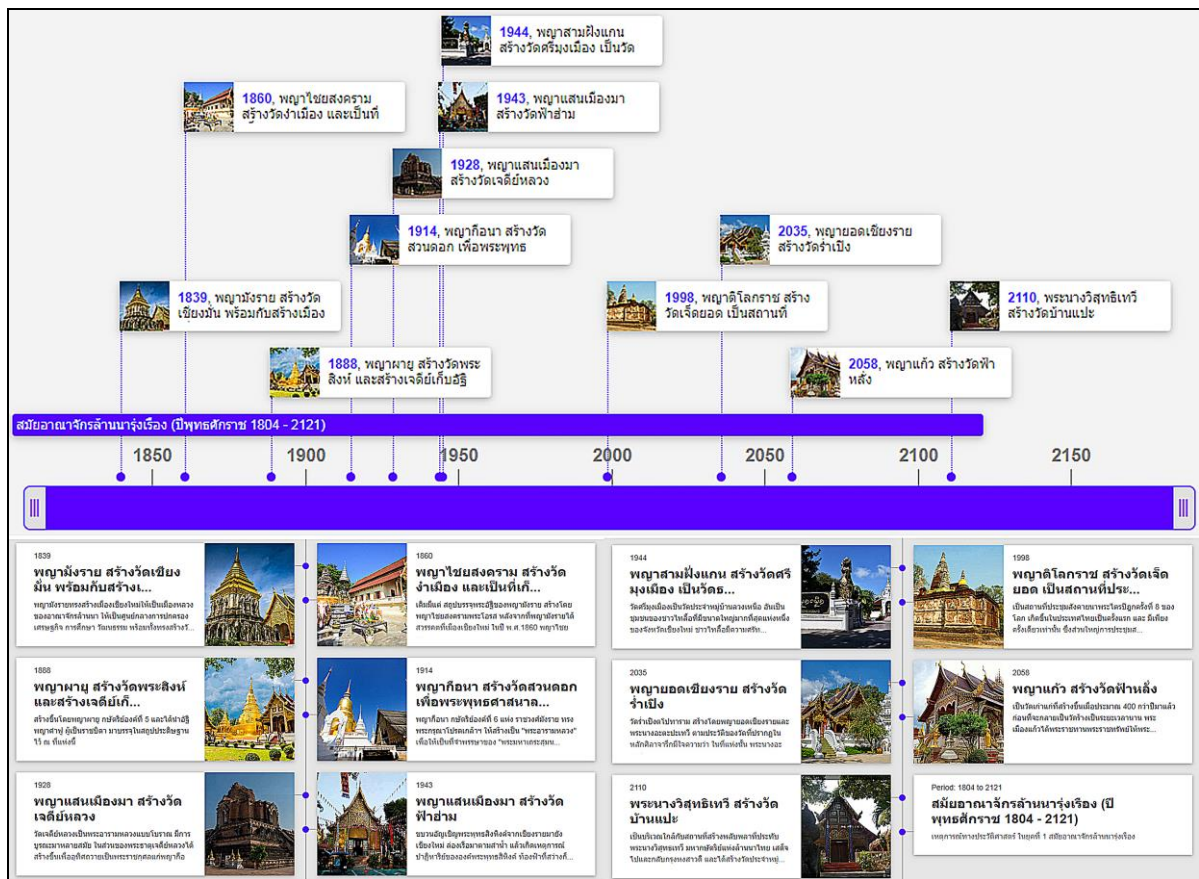
Fig. 9: Web-based applications for Lanna history: chronological search related to the Lanna kingdom

Conclusion

This paper proposes the history relation extraction of Buddhist temple documents, by using the names of Lanna kings as named entities to represent the history timeline.

The three historical relation extraction methods are (1) R1: the name of the king who built the temple (2) R2: the name of the king who created or restored a historical site within temple grounds and (3) R3: the year of temple construction. The process presented here using a Thai word segmentation toolkit and our algorithm gives high accuracy results on three proposed historical relations, in which the extractions can be further used as knowledge-based information of the Lanna kingdom.

This process can be extended to be used as a summary generator of Buddhist temple documents based on these three relations.

However, the algorithm needs to be improved to overcome the previously discussed problems, in order to be used on other temple documents which covers Thai kings from other dynasties. With the improved algorithm we will be able to extract further historical relations apart from R1, R2 and R3, which will assist history experts in their studies.

For future research, we will improve development and extension of our approach to be applied in other areas of historical domains such as important historic events, culture and agriculture. Moreover, we plan to apply timeline visualization techniques to present historical events with accurate results.

Acknowledgement

Author's Contributions

Pongkorn Chantaraj: Data Collection, system design and development the method for extract the historical relations from Buddhist temple documents and writing and revising the article.

Jaratsri Rungrattanaubol: Participation in discussions on the methods, the organization of ideas and writing and revising the article.

Anamai Na-udom: The organization of the article, writing and revising the article.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Aboaoga, M. and M.J.A. Aziz, 2013. Arabic person names recognition by using a rule based approach. J. Comput. Sci., 9: 922-927.
DOI: 10.3844/jcssp.2013.922.927

Abraham, S., S. Mäs and L. Bernard, 2018. Extraction of spatio-temporal data about historical events from text documents. Trans. GIS, 22: 677-696.
DOI: 10.1111/tgis.12448

Aunanan, A., S. Kankaew, N. Wannapirun and T. Khongsomlap, 2016. The Thai news analysis to auto tagging using text extraction. Proceedings of the Lecture notes in 2016 1st Rajamangala University of Technology Suvarnnabhumi National Conference, Jun. 22, IEEE Xplore Press, Thailand, pp: 427-434.

Baker, C. and P. Phongpaichit, 2014. A History of Thailand. 3rd Edn., Cambridge University Press. ISBN-10: 978-1-107-42021-2.

Biller, O., A. Asi, K. Kedem, J. El-Sana and I. Dinstein, 2013. WebGT: An interactive web-based system for historical document ground truth generation. Proceedings of the 12th International Conference on Document Analysis and Recognition, Aug. 25-28, IEEE Xplore Press, Washington, DC, USA, pp: 305-308. DOI: 10.1109/ICDAR.2013.68

Bovornkitti, L., 2005. The influence of buddhism in historical Thai art. J. Royal Institute Thailand, 30: 355-362.

Chen, Y.J. and J.Y.J. Hsu, 2016. Chinese relation extraction by multiple instance learning. Proceedings of the Workshops at the 13th AAAI Conference on Artificial Intelligence: Knowledge Extraction from Text, Feb. 12-13, Phoenix, Arizona USA, pp: 502-506.

Cybulska, A. and P. Vossen, 2011. Historical event extraction from text. Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Jun. 24, IEEE Xplore Press, Portland, pp: 39-43.

Hao, Q., J. Keppens and O. Rodrigues, 2017. A verb-based algorithm for multiple-relation extraction from single sentences. Proceedings of the International Conference on Information and Knowledge Engineering (IKE), the Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing, pp: 115-121.

Haruechaiyasak, C. and A. Kongthon, 2013. LexToPlus: A Thai lexeme tokenization and normalization tool. Proceedings of the 4th Workshop on South and Southeast Asian Natural Language Processing, Oct. 14-18, Nagoya, Japan, pp: 9-16.

Hasan, K.T., S.S. Abdullah, R. Ahmed and F. Giunchiglia, 2013. The history of temporal data visualization and a proposed event centric timeline visualization model. Int. J. Comput. Applic., 70: 27-33. DOI: 10.5120/12241-8497

Kauppinen, T., J. Väätäinen and E. Hyvönen, 2008. Creating and using geospatial ontology time series in a semantic cultural heritage portal. Proceedings 5th European Semantic Web Conference the Semantic Web: Research and Applications, Jun. 1-5, IEEE Xplore Press, Tenerife, Canary Islands, Spain, pp: 110-123. DOI: 10.1007/978-3-540-68234-9_11

Kirigaya, K., 2014. Some annotations to the Chiang Mai chronicle: The Era of Burmese Rule in Lan Na. J. Siam Society, 102: 257-290.

Mooney, R.J. and R. Bunescu, 2005. Mining knowledge from text using information extraction. ACM SIGKDD Explorations Newsletter, 7: 3-10. DOI: 10.1145/1089815.1089817

Nakatoh, T., C. Yin and S. Hirokawa, 2011. Extraction and disambiguation of name of place from tourism blogs. Proceedings of the 1st ACIS International Symposium on Software and Network Engineering, Dec. 19-20, IEEE Xplore Press, Seoul, South Korea, pp: 73-78. DOI: 10.1109/SSNE.2011.29

Ongsakul, S., 2005. History of Lanna. Silkworm Books. ISBN-10: 9789749575840.

Penth, H., 2001. A brief history of Lanna: Northern Thailand from Past to Present. Silkworm Books. ISBN: 9781628409505.

Piyabhani, P.N., 2016. Sukhothai Kingdom: The Golden Age of Buddhism. J. Int. Buddhist Studies, 7: 123-131.

Saif, A.M. and M.J.A. Aziz, 2011. An automatic collocation extraction from Arabic corpus. J. Compu. Sci., 7: 6-11. DOI: 10.3844/jcssp.2011.6.11

Sazali, S.S., N.A. Rahman and Z.A. Bakar, 2016. Information extraction: Evaluating named entity recognition from classical Malay documents. Proceedings of the 3rd International Conference on Information Retrieval and Knowledge Management, Aug. 23-24, IEEE Xplore Press, Bandar Hilir, Malaysia, pp: 48-53. DOI: 10.1109/INFRKM.2016.7806333

Sharum, M.Y., M.T. Abdullah, M.N. Sulaiman, M.A.A. Murad and Z.A.Z. Hamzah, 2011. Name extraction for unstructured Malay text. Proceedings of the IEEE Symposium on Computers and Informatics, Mar. 20-23, IEEE Xplore Press, Kuala Lumpur, Malaysia, pp: 787-791. DOI: 10.1109/ISCI.2011.5959017

Sutheebanjard, P. and W. Premchaiswadi, 2009. Thai personal named entity extraction without using word segmentation or POS tagging. Proceedings of the 8th International Symposium on Natural Language Processing, Oct. 20-22, IEEE Xplore Press, Bangkok, Thailand, pp: 221-226. DOI: 10.1109/SNLP.2009.5340914

Tepdang, S., C. Haruechaiyasak and R. Kongkachandra, 2010. Improving Thai word segmentation with named entity recognition. Proceedings of the 10th International Symposium on Communications and Information Technologies, Oct. 26-29, IEEE Xplore Press, Tokyo, Japan, pp: 940-945. DOI: 10.1109/ISCIT.2010.5665124

TIC, 2016. The data center of the temples in Thailand. http://www.templethailand.org

Tongtep, N. and T. Theeramunkong, 2009. A feature-based approach for relation extraction from Thai news documents. Proceedings of the Pacific-Asia Workshop on Intelligence and Security Informatics, Springer Berlin Heidelberg, pp: 149-154. DOI: 10.1007/978-3-642-01393-5_16

Wikaha, P. and P. Netisopakul, 2014. A study of named entity recognition. KMITL J. Inform. Technol., 3: 32-37.

Wonglangka, W. and F. Han, 2018. Lanna buddhist temple: The cultural landscape as urban green space. J. Environ. Design, 5: 107-129.

Zayed, O.H. and S.R. El-Beltagy, 2012. Person name extraction from modern standard Arabic or colloquial text. Proceedings of the 8th International Conference on Informatics and Systems, May 14-16, IEEE Xplore Press, Cairo, Egypt, pp: 44-48.