Constructing a Rule Based Naming System for Thai Names Using the Concept of Ontologies

Chakkrit Snae
Department of Computer Science
Naresuan University
Phitsanulok, Thailand
chakkrits@nu.ac.th

Abstract

Names are important in many societies, even in technologically oriented ones which use ID systems or other ways to identify individual people. Names such as personal surnames are the most important as they are used in many processes, such as identifying of people, record linkage and for genealogical research as well. For Thai names this situation is a bit different in that in Thai the first names are most important. Even phone books and directories are sorted according to the first names. Here we present a system for constructing Thai names from basic syllables. Typically Thai names convey a meaning. For this we use an ontology of names to capture the meaning of the variants which are based on the Thai naming methodology and rules.

1 Introduction

Names are used for identifying persons, places, things and even ideas or concepts. Names serve for labelling of categories or classes and for individual items. They are properties of individuals which are of greater importance in most communities. In technological oriented societies such as modern Western the reference between names as a label and the person is not as obvious as in small tribal societies. This is especially true where names are stored within large information systems. This includes government, medical, educational and even commercial records which are kept about individuals. Names are the most important referrer to a person even if there are numbering systems like ID numbers because such systems are not universal. Names are often queried in a different way than they were entered. Names represent complex lexical structures which have to be handled systematically for data entry, storage and retrieval in order to get sufficient recall or precision the retrieval process in.

In this paper we present a first account of our findings on constructing Thai names with the help of an ontology of names as well as a working methodology for the naming process in Thai culture.

This paper is organized as follows: Section 2 contains a description of names and their elements. In Section 3 we outline the concept of ontology of names. In Section 4 we present the construction process and system for Thai names. We apply this system together with an ontology to construct names with an appropriate meaning. Section 5 shows the conclusions of our study and further work which has to be performed.

2 What is a Name?

Names for individuals are often called proper names, for humans sometimes also anthroponyms. Names for places are called toponyms, for bodies of water hydronyms, for ethnic groups ethnonyms, for metaphors metonyms and so on. Names are more than just strings of characters. Names show important information such as titles, gender, marital status, and even birthplace. For this names provide different elements (Section 2.2), which may differ between cultures.
2.1 Naming for Identity and Security

From the technical point of view we want to link and match as many names as possible with the correct individuals. If we deal with individuals of the same name, e.g. John Smith, we have to establish a second identifier at least. This can be – and is in many cases – a temporal element, like the date of birth, which is an individual and unchanging property of the person. Another way to circumvent the problem is to establish numbering systems, like ID numbers. Systems of numbers or other ciphers can be generated within individual organisations. It is not likely that the resulting ID numbers will be the same in different organisations. The numbering may have limitations as well, e.g. the individual health care setting (e.g. within a hospital or district) or, in principle, more widely (e.g. the National Health Service number). In the past, the National Health Service number in England and Wales had serious limitations as a matching variable, and it was not widely used on health-care records. With the allocation of the new ten-digit number throughout the NHS all this has been changed (Gill, 1997).

Although numbering systems are simple to implement they can lead to different errors in recording, transcription, and keying. So we have to take into account methods which reduce these errors and facilitate good quality of data entry and retrieval. One such method uses a checking device such as check-digits (Wild, 1968, Hamming, 1986). When we are not able to use unique numbers or ciphers, natural matching variables are the person’s name, date of birth, sex and perhaps other supplementary variables such as the address with postal code and place of birth, which are used in combination for matching. Recently, it has been suggested that this simple code could be extended for security critical places (e.g. airports, checkpoints etc.) with biometric marker information extracted from person identifier information e.g. fingerprints/iridograms.

2.2 Elements of Personal Names

The following table shows typical elements of personal names together with potential variations and sources of choices, e.g. dictionary of given names.

| 1. Initial and familial names                              | 2. First name, given name                        |
|------------------------------------------------------------|-------------------------------------------------|
| • for male - Mr                                           | • number of parent names                        |
| • for female - Miss                                       | • number of grandparents                       |
| • single female - Ms                                       | • grandparent                                  |
| • married female - Mrs                                    | • using name that matches character             |
| • professional Occupation - Dr., Prof.                   | • e.g. Mr, Mão                                  |
| • honorary titles and familial titles for formal status in society - Lord, Sir, Knight, Baron, etc. | 3. Middle name                                  |
| • using name that matches character                      | 4. Surname                                     |
| • e.g. Mr, Mão                                            | • inherited from family                        |
| • using name that matches character                       | • the king or Royal family                     |
| • e.g. Mr, Mão                                            | • parent surnames                              |
| • using name that matches character                       | • grandparents                                 |
| • e.g. Mr, Mão                                            | • using name that matches character             |
| • e.g. Mr, Mão                                            | • e.g. Mr, Mão                                  |
| • using name that matches character                       | • called by family                             |
| • e.g. Mr, Mão                                            | • called by second name                        |
| • using name that matches character                       | • called by third name                         |
| • e.g. Mr, Mão                                            | 5. Nicknames                                   |
| • using name that matches character                       | • last syllable of given name, e.g.            |
| • e.g. Mr, Mão                                            | • first syllable of given name, e.g.           |
| • using name that matches character                       | • Robert ≡ Bob                                 |
| • e.g. Mr, Mão                                            | • called by family                             |
| • using name that matches character                       | • called by second name                        |
| • e.g. Mr, Mão                                            | 6. Artist name and pseudonyms                  |
| • using name that matches character                       | • first name only, like Sasha                  |
| • e.g. Mr, Mão                                            | • surname only, like Ching możemy                 |
| • using name that matches character                       | • only nickname, like Prince                    |
| • e.g. Mr, Mão                                            | • only abbreviation, like H.L.                 |
| • using name that matches character                       | • for artist, model, etc.                      |

Figure 1: The elements of names

3 Ontology of Names?

The term ontology has been widely used in recent years in the field of Artificial Intelligence, computer and information science especially in domains such as, cooperative information systems, intelligent information integration, information retrieval and extraction, knowledge representation, and database management systems (Guarino, 1998, Andrade and Saltz, 1999, 2000). Many different definitions of the term are proposed. One of the most widely quoted and well-known definition of ontology is Gruber’s (Gruber, 1993): An ontology is an explicit specification of a conceptualization.

The term is borrowed from philosophy, where an ontology is a systematic account of existence. Here in this paper we adopt the following definition: Ontology is the study or concern about what kinds of things exist - what entities or things are there in the universe (Blackburn, 1996). Our work on ontologies will comprise: a terminological component where we lay down the concepts and an assertional component (or Knowledge Base) which contains the individual instances (entities). The level of description will be taxonomies with hierarchically related terms and controlled vocabularies (thesaurus) with the help of semantic networks.

An ontology of names can be worked out in many different forms, but every ontology will include a dictionary, some definition of the terms.
(semantics), and indications how they depend on each other, e.g. in hierarchies and semantic networks. For example, an ontology of names can be defined as what kinds of names exist, e.g. first name, surname, nickname, etc (Section 2.2). This typically comprises definitions of different names, the elements of names and their structures. In this section we show how an ontology of names can be captured and defined.

An ontology can also be used to establish the network of synonyms, e.g. using spelling norms to determine whether two names are the same/similar or not. For example, two names: Totie and Totiey can be defined based on assumption that they are the same as Totty. This attempts to tackle the seemingly irreducible conventions of surname. In compositional semantics let us consider the name “Gutensohn”. This name will be used to illustrate the various semantic considerations in German naming. The name is a composition of the two strings Godith and Sohn, which have unambiguous, meaningful interpretations. The interpretation of Godith is god or good battle and Sohn is interpreted as a male child in relation to his parent. The composition Gutensohn, Gudzon, or in other cultures: Guditson, Godye- son and Godithson and Godison (Reaney and Wilson 1997).

We incorporate the different elements of personal names (Figure 1) into a semantic network (Figure 2) to illustrate how they associate with each other, e.g. with hierarchies.

Figure 2: Representation of the names elements using semantic nets
Identifying and searching result in a list many names with variations and meanings. In order to find the correct person with a name we have to adopt ontologies of names, e.g. based on place of birth or relationship of people. The typical origins of surnames which can be a basis for ontologies of names can be classified as follows:

- **Local surnames** - surnames of relationship - surnames of occupation or office.

Local surnames, which are most widely used, stem from toponyms, we can call them toponymic. They reflect land owners, place of birth, or the center of life. For example, Richard de Tonebridge was named after his castle of Tonbridge, but he was also called Richard de Clara from the Suffolk Clare, which became his chief seat and the family’s definitive surname. Also Richard de Hadestoke, a London alderman, had left Hadstock (Essex) and settled in London (Reaney and Wilson 1997). These local surnames derive (with occasional exceptions) from English, Scottish or French places (e.g. de, at, in). Toponymic Thai names are derived from Thai places and took originally a preposition na, for example, Prapas na Ranong is a person from a Southern province in Thailand called Ranong.

Surnames which come from family relation are often called patronymic, but we have to introduce a more elaborate term, because we encounter names from females and other relations than just father, such as Gilbert Fathevedeppeson, Richard Hannebrothir, America Ibbotdoghter, and John Prestebruther.

Surnames of occupation and office refer to actual office holders like clergy names or state offices. Some of these, such as steward, constable, marshal, etc., became hereditary and gave rise to hereditary surnames, but the terms were also commonly used of lesser offices, whilst marshal was a common term for a farther and such names frequently denoted the actual occupation. However, Nuns, Abbots, Priors, Monks and other clerical people were bound by vows of celibacy and thus did usually not have families which adopted their respective surname.

4 Rule Based Naming System for Thai Names

In this section we introduce the well known methodology for Thai naming process as well as how Thai names can be constructed using the basic Thai rules of forming syllables.

4.1 Custom Naming Process Using Thai Astrology

The way of naming can vary, e.g. naming by monks, grandparents. Since former times names are very important to people. Naming from the past to the present has been continuously developed and has developed a variety of patterns. Each pattern has it own rules depending on local places and the belief that has been developed until the present. The basic goal of naming is to provide a good fortune and progress during life. Most first names have a meaning. The widely used methodology of Thai naming process is briefly described in the following.

Principal naming using Thai astrology is widely used since the past. Because it uses the birth day to form the name. This is a belief that the individual has a set of 8 attributes called name of the angles referred to in Thai astrology. These attributes influence each person’s livelihood, fortune, etc. The attributes refer to Servant >Age> Power> Honour> Property> Diligence> Patron> Misfortune. Each attribute has it own letters which can be used for constructing names.

4.2 Syllable Construction

Syllables are aggregated to names which sound good or aimed at good fortune according to the methodology mentioned above. As a consonant can not stand alone in Thai language and personal names we consider rules for vowels only. The order is:

- Vowels can come first or can be followed by a first consonant, e.g. Ek
- Vowels can follow a first consonant without a final consonant, e.g. Ka
- Vowels that can not have final consonant, e.g. Tam, Tua
- Vowels that need final consonant, e.g. Kak

![Figure 3: Forming of Thai syllables](image)
Example of construction of Thai syllables using Thai Romanization 1.10 unicode (CU 2004) according to Figure 3: \( n \) (Ka) = CV, \( \upsilon \) (Ek) = VC, \( \eta \) (Kok) = CF, \( \hat{n} \) (Kak) = CVF, \( \upsilon \eta \) (Ek) = VF.

Thai names are built from one or more syllables that may or may not have a meaning. There are syllables which alone do not mean much in particular, but when used as prefixes and suffixes can change the meaning of the syllables they precede or follow as stated below.

Prefixes: for example, (1) kob means "gather" while pra-kob means "put together" or "consist of", (2) cham means "remember" while pra-cham means "regularly".

Suffixes: for example, (1) ngarm means "beautiful" and num means "water". Add jai ("heart") to each and we have names like ngarm-jai and num-jai meaning "beautiful in the heart" and "generous" respectively.

In the following it is shown how to construct Thai names that convey a meaning with the help of ontologies. Syllables are built from consonants (either C or F, C being the first and F the final consonant) and vowels. A name consists of one or more syllables. One syllable can have a meaning of its own, which leads in case of two or more syllables in a name to more complex meanings.

The process of constructing names according to the naming rules and methodology begins with a leading consonant or vowel that can be the only letter in the name. If we continue to add more letters we come either to a valid name (a name which has a meaning) or to an invalid name (a name without a meaning). Invalid names will be discarded in such a way that the last letter will be replaced by another or will be added with more letters.

In Figure 4 it is shown that names comprise \( n \) syllables with a reasonable number of letters. The meanings of the syllables as well as of the name are found with the help of an ontology of names.

The meaning of the name for a girl Pensri in the example (see Figure 4) is “the goodness and beauty of the moon”.

How do we know which name belongs to a boy or a girl? There are several ideas to take into consideration when selecting a name for indicating the gender. Ontologies can help in deciding the name by capturing and indexing the meaning as is shown in Table 1.

Boys can be named by:
- taking names of leading or important male characters accepted as having lasting literary value, e.g. Chakkri (an actor from KhunChang KhunPhaen) and Suthat (an actor from Phra Aphai Mani).
- combining monosyllables evidently indicating the males sex, e.g. Chatri (a mighty man), Choetchai (a perfect man), and Danai (son).
- using adjectives indicating male qualities such as strength, and bravery, e.g. Watcharaphon (strength), woraphon (strength), and Kriangkrai (bravery).
- using terms representing something strong, important or high, e.g. Suriya, Phassakorn, and Phanuphong are popular names. They all mean the sun.

Choosing names for girls is even more complicated. There are so many things to consider (Table 2). A girl can be named by:
- taking name of leading female characters from well-known literature: Phim Phi La Lai (an actress from KhunChang KhunPhaen) and Suwanmali (an actress from Phra Aphai Mani).
- combining monosyllables evidently indicating the females sex, e.g. Charun (pretty girl), Kanyarat (pretty woman), Khantida (younger sister), and Khwansuda (beloved daughter).
- using syllables indicating female qualities such as those having to do with looks and beauty, e.g. Pariyaphat (lovely) and Phichittra (beautiful).
• using syllables describing softness, gentleness and mildness, e.g. Pranit (gentle), Lamun (soft), and Sukhum (mild).

• using syllables describing scent, taste, e.g. Saowakhon (lovely smell, fragrant) and Parimon (fragrant), Mathurot (sweet) and Wasita (sweet and aromatic). On the other hand, unfavourable tastes like Khom (bitter) or Fad (sappy) are not used.

• using syllables which are names of ornaments and jewellery, e.g. Phatchara Walai (diamond bracelet), Rattana Wali (gem necklace), Phara (diamond), and Rachawadi (blue gems).

• using syllables which are names of flowers, flowering trees: Manlika (jasmine), Maliwan (jasmine), and Watsana (Charisma), all of which are names of flowers found in Thailand; also syllables which simply mean flower: Buppha, Phakamat, Butsabong, and Ladawan.

| Thai name | Romanized | Meaning | Cluster |
|-----------|-----------|---------|---------|
| ขันดี | Choktin | an actor from KhunChiang KhunPhaen | leading or important male characters from well-known literature |
| จักรกลุกแอร์ | Chek.Krit | an actor from Unnarut | |
| สุทธา | Suchet | an actor from Phra Aphai Mani | |
| สันติชัย | San Chai | an actor from WatSawang Chaidok | |
| อิทธิพล | InThornChat | an actor from Rammakien | |
| ธาตุรื่น | Chatri | a mighty man | terms indicating the males sex |
| สินตานต์ | Chooet Chai | a perfect man | |
| จักรดี | Yor Chai | a brilliant man | |
| นพพิทย์ | Nophphon | a prestigious and vigorous man | |
| วิศวพันธ์ | Wi Chon Phon | a powerful man | |
| ทนัย | Donrai | a son | |
| ไพฑูรย์ | Watchara Phon | strength | adjectives indicating male qualities such as strength, and bravery |
| นรทล | Wonphon | strength | |
| พลประทีป | Phis Thip | strength | |
| กิจฤทธิราช | Hrangtrai | bravery | |
| ปิยพล | Wise Phat | bravery | |
| ทวีสิริ | Thira Sen | brave and quiet | |
| สุริยา | Suriya | the sun | terms representing something strong, important or high |
| ภราภักดี | Photaphak | the sun | |
| ภานุพงษ์ | Phanup Phong | tribe of the sun | |
| กัสสันต์ | Rachan | the king | |
| อริยวัชร | Hruison | lion king | |
| ประสิทธิ์ | Chet Chai | triumph of the king | |
| ชนะชัย | Chana Thip | the king | |

Table 1 Examples of Thai names with their meanings according to male gender.
Table 2 Examples of Thai names with their meanings and female gender.

This grouping process is used to build an ontology which has a (classified) database structure. This allows for speeding up the information retrieval process for the naming system.

4.3 Web Based Naming Expert System

Currently we are constructing and implementing a web-based naming expert system which offers two basic ways to come to “good” Thai names according to the first methodology mentioned above. The system will give us the letters for each date of birth. We use these letters to construct names based on the basic rules (see Figure 2). The user will be able to choose from a list of resulting possible names according to their respective meaning.

We use a dictionary database of more than 8,000 Thai names which contains not only the spelling, but also the meaning and correct pro-
nunciation. In situations where names follow the rules but do not have a meaning we compare the name with similar names in a dictionary database and check for similarity using a simple string matching scheme. Then the user can select the best name from the resulting list of names.

A second way to come to names is by using ontologies instead of basic spelling rules which are used according to the sex and date of birth. For this we check the different names against the date of birth by implementing an indexed database system of names from Thai dictionary for every day of a week.

5 Conclusion and Further Work

We have used Thai customs for the naming process, an ontology of names, and the basic rules for forming syllables in Thai to construct the rule based naming system. We want to extend the system using name matching algorithms to return the variants of names from a Thai dictionary with the relative probability of their similarity. To speed up this process we will use a database structure based on an ontology of names, e.g. by indexing names according to gender and meaning with the help of a Thai dictionary database. We will use a Romanized version of Thai names.

The advantage of this process would be an improvement of searching algorithms for Thai names in databases as well as in the internet. Here we will need name matching algorithms. The next step of development is to take into account different cultures.

Currently we are designing a system for multicultural name matching called NARESUAN-M² (NAme Recognition Expert System Using Automated Name Matching Methods). A primary objective here is to study how ontologies and algorithms can help in deciding which rules of naming system have to be implemented. This will also require an investigation into how ontologies that cover the different elements of names can be merged.

References

Henrique Andrade and Joel Saltz. 1999. Towards a Knowledge Base Management System KBMS: An Ontology-Aware Database Management System DBMS, Proceedings of the 14th Brazilian Symposium on Databases, Florianopolis, Brazil.

Henrique Andrade and Joel Saltz. 2000. Query Optimization in Kess – An Ontology-Based KBMS. Proceedings of the 15th Brazilian Symposium on Databases (SBBD’2000). João Pessoa, Brazil.

Simon Blackburn. 1996. The Oxford Dictionary of Philosophy, Oxford: OUP.

CU.2004. <http://www.arts.chula.ac.th/%7Eling/tts/> (Nov. 19, 2005)

Kerry Dematteis, Richard Lutz and Heather McCallum-Bayliss. 1998. Whose Name Is It: Names, Ownership and Databases. Originally written for: 1998 Annual Meeting American Name Society San Francisco, CA.

Leicester E. Gill. 1997. OX-LINK: The Oxford Medical Record Linkage System, Complex linkage made easy, Record Linkage Techniques. In: Proceedings of an International Workshop and Exposition, 15-33.

Thomas R. Gruber. 1993. A Translation Approach to Portable Ontology Specification, Knowledge Acquisition, Vol. 5, No. 2, 199-220.

Nicola Guarino. 1998. Formal Ontology and Information Systems. In: Nicola Guarino (Ed.): Proceedings FOIS’98, Amsterdam.

Richard W. Hamming. 1986. Coding and Information Theory. 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Daniel S. Jurafsky and James H. Martin. 2000. Speech and Language Processing, Prentice Hall.

Percy H. Reaney and R.M. Wilson. 1997. A Dictionary of English Surnames, Oxford: OUP.

W.G. Wild. 1968. The Theory of Modulus N Check Digit Systems. The Computer Bulletin, 12, 308-311.

Wikipedia. 2005. <http://en.wikipedia.org/wiki/Royal_Thai_General_System_of_Transcription> (Nov. 19, 2005)

Ian Winchester. 1973. On referring to ordinary historical persons. In: Identifying People in the Past. E.A. Wrigley (Ed.), 17-40.

Ian Winchester. 1970. The Linkage of Historical Records by Man and Computer: Techniques and Problems. Journal of Interdisciplinary History. 1, 107-124.