Asian Interdisciplinary and Sustainability Review

Original Research Article

Received: 10 October 2023 **Revised:** 23 January 2024 **Accepted:** 23 January 2024

DEVELOPMENT OF DATA CLASSIFICATION OF LOCAL HERBS IN BETONG, THAILAND

Kwanrutai SAELIM[1], Suppamas RATTANAPIPAT[1], Muhammad PU[1], Amart SULONG[1], Sothana VICHARUEANG[2] and Lakjaree SPHANURATTANA[2]

1 Faculty of Management Sciences, Yala Rajabhat University, Thailand; kwanrutai.s@yru.ac.th (K. S.)

2 Faculty of Science and Technology, Thammasat University, Thailand

Handling Editor:

Professor Dr.Roy Rillera MARZO Management and Science University, Malaysia

(This article belongs to the Theme 2: Science and Technology for Sustainable Development)

Reviewers:

1) Associate Professor Dr.Pannee SUANPANG Suan Dusit University, Thailand
2) Dr.Attaporn WANGPOONSARP Prince of Songkla University, Thailand
3) Dr.Nuanluk SANGPERM Kasetsart University, Thailand

Abstract

This research focuses on the development of an application named "Piyamit Herb," dedicated to preserving herbal information in the Thai language within Betong, Yala, Thailand. The primary objectives of the application are to digitize information previously conveyed verbally within the Piyamit Community in Betong, ensuring its accessibility to interested parties and the younger generation. The application also serves to systematically store this information, alleviating concerns about its potential loss over time. Additionally, it aims to foster a sense of ownership of local learning resources within the community. The Piyamit Community boasts a rich repository of indigenous herbs used in traditional medicine. Consequently, there is a need for classifying these local herbs to facilitate learning and usage. The development of the application incorporates ontology, a technology that facilitates the linking of lexical search and categorization of information. Furthermore, Natural Language Processing theory in the Thai language, text classification, and machine learning technologies are applied. As a result, Piyamit Herb has been successfully developed as a comprehensive system for storing and presenting data in Betong, Thailand. The application addresses the imperative to preserve information shared orally by local elders, ensuring the transmission of valuable knowledge passed down from their ancestors.

Keywords: Natural Language Processing, Machine Learning, Data Classification, Keywords Extraction

Citation Information: Saelim, K., Rattanapipat, S., Pu, M., Sulong, A., Vicharueang, S., & Sphanurattana, L. (2024). Development of Data Classification of Local Herbs in Betong, Thailand. *Asian Interdisciplinary and Sustainability Review*, *13*(1), 12-23. https://doi.org/10.14456/aisr.2024.2

Introduction

Nestled within the enchanting Betong in the heart of the City of Yala, Thailand lies the Ban Piyamit community, a bastion of cultural richness and biodiversity. Amidst the verdant landscapes, the invaluable knowledge encapsulated in local herbs thrives, serving as a testament to the deep-rooted connection between nature and the Ban Piyamit community. In recognition of the significance of preserving this botanical heritage, this article embarks on a purposeful journey — the development of a comprehensive data classification system for the local herbs of Ban Piyamit. Beyond the ethereal beauty of the surroundings, the need to create a robust database highlighting the importance of specific herbs in this community area becomes paramount. As we delve into the intricate tapestry of traditional knowledge and modern technology, the creation of a database emerges as a pivotal aspect of this initiative. Our focus extends beyond mere documentation; it is a commitment to empower the Ban Piyamit community by cataloging the essential herbs that have played an integral role in their cultural practices and holistic well-being. Join us in this exploration as we navigate the landscapes of Betong, acknowledging the need to create a comprehensive database that not only preserves the wisdom of Ban Piyamit but also paves the way for sustainable practices and community empowerment. Through this endeavor, we aim to weave together the threads of tradition and innovation, fostering a harmonious coexistence between nature and the Ban Piyamit community.

The valuable information of local herbs and their medicinal treatments in Betong in the area of Yala has always been passed on from generation to generation verbally as a traditional way of how people pass on information previously. However, this method seems to be rather ineffective because the passed information or knowledge can be missing or ambiguous due to personal capability and understanding. We have been thinking that it must be better to develop a more practical method to collect and store this beneficial knowledge accurately and precisely to be able to pass it on to our future generations especially young people in the local area and to be able to exchange this knowledge with other areas in Thailand. Therefore, we have developed an application called Piyamit Herb to classify local herb information in Betong and store it in a format which can be easily displayed and shared publicly online. We have extracted keywords or key phrases to process textual data of local herbs. The challenging points of this project are to process complex descriptions of each herb in the Thai language and to classify herbs according to their treatment properties to reduce the complexity of the information we have. We also aim to find the relation-ships among local herbs through keyword extraction. Ganoderma lucidum is a versatile and valuable mushroom with a wide range of nutritional, cosmeceutical, mycochemical, pharmacological, and clinical properties. The review underscores the need for further research and standardization to fully harness its therapeutic potential. This research provides a comprehensive overview of Reishi mushroom, covering its multiple dimensions from nutrition and cosmetics to pharmacology and clinical applications (Ahmad et al., 2021). Ganoderic acids may have cardiovascular benefits by reducing blood pressure, improving lipid profiles, and protecting against cardiovascular diseases. The review discusses the hepatoprotective properties of Ganoderma triterpenoids, including their ability to protect the liver from damage and promote liver health. The paper highlights the diverse molecular mechanisms of Ganoderma lucidum triterpenoids, which contribute to their potential health benefits. Understanding these mechanisms is essential for exploring the therapeutic applications of Reishi triterpenoids (Liang et al., 2019).

Literature Review

This application primarily employs two techniques for classifying local herbal information in Betong. The challenges and pain points in this application stem from the nature of the Thai language and the previous lack of a standardized data format for storing local herbal information. Due to the absence of an official format, this valuable information has been transmitted verbally from one generation to the next, leading to inaccuracies and imprecisions in its transfer. Moreover, the reliance on verbal transmission poses the risk of losing this knowledge if there are no willing individuals to learn and pass it on in the future. These factors underscore the necessity of developing the Piyamit Herb application, which aims to formally learn, store, and transfer all information and knowledge, ensuring its preservation for future generations and facilitating distribution or exchange with people from other areas in Thailand.

To make Piyamit Herb effective for a broad audience, we have implemented two methods: keyword extraction and text classification. Keyword extraction is utilized to identify the main words or phrases associated with each herb, while text classification is employed to group herbs based on their healing properties. The major challenge in our work lies in dealing with the Thai language, known as one of the most challenging languages to process automatically.

Piyamit Village is a community of Chinese descendants residing in Betong, Yala, Thailand. These Chinese individuals migrated to the southern border areas of Thailand as part of the Communist Malaya Party, which fought against Japanese soldiers occupying the region in 1942. Despite the end of World War II, they continued to inhabit the same area, now jungles, refusing to join the British colonies ruling the Malaysian government. During this period, the Communist Malaya Party, unable to venture out of the jungles for supplies, relied solely on wild herbs for medicinal purposes. After the war, this community furthered their knowledge of herbs and their treatments, even traveling to China to explore indigenous herbal medicine. This expertise led to the development of three initial herbal products: Reishi mushroom, wild pepper root, and Daxi Wang. Reishi mushroom, also known as Ganoderma lucidum, is effective in treating high blood pressure, diabetes, and high lipids. Wild pepper root possesses medicinal properties for treating gout, gallbladder issues, and preventing cancer. Daxi Wang is known for its ability to cure liver disease, jaundice, nourish the brain, and boost energy levels. Additionally, Piyamit Village boasts special identity products such as wild stator root and Wu Zhao Ma Te (five-clawed dragon), each with unique medicinal properties.

Recognizing the importance of preserving the knowledge of local herbs mentioned above for the people of Betong, our inspiration is to systematically store this information. The main objectives are to pass on this regional herbal knowledge to future generations and leverage present technologies to categorize herbs with similar medicinal properties into cohesive groups.

Keywords Extraction

Keyword extraction (KE) is a text analysis technique that automatically identifies the most frequently used and significant terms and phrases within a given text. These extracted words typically best encapsulate the subject matter of the text, aiding in content summarization and the identification of the main ideas (Bharti et al., 2017; Nasar et al., 2019). Automatic keyword extraction (AKE) involves the automated extraction of keywords through computational processes (Bharti et al., 2017; Nasar et al., 2019). In the contemporary landscape, machine learning, particularly artificial intelligence (AI) integrated with natural language processing (NLP), is harnessed to facilitate the understanding and analysis of human language. AKE plays a crucial role in providing a quick overview of textual data. Keywords can manifest as single words or in the form of phrases or multi-word expressions. The preference for extracting key phrases or multi-word terms, as opposed to single words, arises from the potential for misunderstanding associated with single-word terms lacking specificity. Utilizing key phrases helps alleviate this issue (Siddiqi & Sharan, 2015). Keywords and key phrases are sourced from diverse text types, including general documents, reports, social media comments, reviews, and more. Extracted terms serve to represent pertinent information embedded within the texts, addressing the imperative to sift through vast information on the internet and extract crucial content from various sources.

Presently, keyword extraction can be categorized into two types: keyword assignment and keyword extraction. The distinction lies in how keywords are derived. Keyword assignment acquires important terms or keywords based on information within a text, even if these keywords may not appear in the text itself. This method closely resembles how humans assign keywords that best describe a text after reading it. Conversely, keyword extraction exclusively identifies important terms present in the text, emphasizing the importance of building an effective database system through systematic data management. Numerous methods are available for keyword extraction, with one popular approach involving the determination of word frequencies in a document. However, this method has a drawback, as the high-frequency extracted keywords may not consistently represent the entirety of the document.

Various research endeavors aim to enhance the effectiveness of keyword extraction, particularly addressing issues where extracted terms may have similar meanings and should be grouped under the same topic. Topic Rank, an algorithm derived from PageRank, tackles this challenge by grouping synonyms under the same topic, considering the co-occurrence of word positions in the document, and ranking the top five highest-scoring keywords (Saelim & Kongkachandra, 2017; Nokkaew & Kongkachandra, 2018). Further refinement of Topic Rank involves considering words that initially appear in the document (Saelim et al., 2019). Fuzzy C-Mean (FCM) is another method employed to group words based on the degree of membership (Anaya et al., 2017). Noteworthy research, such as Ref. (Kurnia et al., 2020), classified comments using Word2Vec and SVM classifiers to categorize them into five different levels of usage. Research of Sharma et al. (2020) grouped social media commentary using Word2Vec factoring principles to reduce data redundancy with Convolutional Neural Network (CNN) into shorter messages. The synergy between keyword extraction and text classification is evident, as good text classification necessitates effective extraction. Finally, the study in Ref. (Jang et al., 2020) presented meaningful information extraction from big data, classifying them into different categories to predict user behavior and emotions. A hybrid model of Bi LSTM and CNN is employed for more accurate classification results in the context of unstructured data.

Thai Natural Language Processing

Natural Language Processing (NLP) is a dynamic blend of art and science, facilitating the extraction of information from textual data for subsequent analysis. At its core, NLP involves fundamental tasks such as defining what constitutes a word and a sentence. The complexity of these tasks is language-dependent, and the challenges vary across different languages. Thai, in particular, presents unique challenges in NLP processing (Tapsai et al., 2019). Unlike English, Thai utilizes spaces between words, and sentence markers are positioned at the end of a sentence, reducing ambiguity compared to other Asian languages like Chinese and Japanese. However, Thai poses additional difficulties due to the absence of word delimiters, making sentence boundary detection problematic. Consequently, subsequent processes in Thai NLP, such as sentence encoding, machine translation, and text summarization, face challenges stemming from the unsolved issue of sentence boundary detection, making it challenging to discern the exact meaning and identify keywords in a Thai sentence.

The Thai language introduces two common problems related to modern vocabulary usage: ambiguous words and words not present in dictionaries, which can complicate NLP tasks. Additionally, Thai lacks symbols for orthographic forms, resulting in words that can be read in different ways. However, algorithms like PyThaiNLP, a Python NLP package designed for the Thai language, address the challenge of word segmentation. PyThaiNLP effectively resolves the disambiguation of word segmentation in sentences containing words with multiple possible readings. This tool has demonstrated proficiency in tackling this prominent issue, a task often performed manually by Thai natives. Consequently, the process of word segmentation in the Thai language necessitates the incorporation of context clues to derive meaningful phrases, transforming it into phrase segmentation rather than simple word segmentation. In this research, we rely on Thai NLP, specifically leveraging PyThaiNLP.

Classification

Classification, a form of supervised learning, involves the utilization of a predefined sample model to allocate sufficient data to respective algorithmic pools. This method ensures specificity, as the system learns based on given inputs and outputs. In the context of this learning paradigm, the programmer, akin to a teacher, assigns appropriate values for specific inputs, aiming to train the system through continuous computation with varying inputs and outputs, thereby establishing meaningful connections. While this process aligns with established principles of learning, a more nuanced examination reveals the role of the programmer as a guiding force in shaping the system's understanding.

An illustrative example of machine learning in classification is found in a research endeavor focused on categorizing user stories (Jurisch et al., 2020). The dataset utilized in this study comprises information from companies engaged in direct exchanges between developers. Given the voluminous nature of data entering the system daily, the challenge lies in training the classification model to adeptly handle this continuous influx of information.

Extending the principles of classification to the realm of herbal information, an effective strategy involves categorization based on medicinal properties. Medicinal herbs sharing the ability to address similar health conditions naturally lend themselves to grouping. Drawing inspiration from the computational principles outlined in Nokkaew & Kongkachandra (2018), we apply a similar methodology to our algorithm. This approach allows for the classification of herbs into distinct classes, offering a systematic framework for organizing and understanding the wealth of herbal information at our disposal.

In conclusion, the literature review has provided a comprehensive exploration of the essential components underpinning the development of the Piyamit Herb application. From the challenges posed by the unique characteristics of the Thai language and the absence of standardized formats for local herbal information to the pivotal roles of natural language processing, keyword extraction, and classification techniques, this overview lays the groundwork for the subsequent discussion. The interplay between linguistic intricacies, computational methodologies, and the rich cultural context of the Ban Piyamit community underscores the significance of our research endeavor. As we navigate the intricate landscape of language processing and knowledge preservation, the synthesis of traditional wisdom and cutting-edge technologies emerges as a driving force. The forthcoming sections will delve into the methodology, implementation, and results, offering a deeper understanding of how the Piyamit Herb application endeavors to bridge the gap between heritage and innovation, contributing to the sustainable preservation and dissemination of invaluable herbal knowledge.

Materials and Methods

Piyamit Herb intricately weaves together three fundamental components, each playing a crucial role in the overarching objective of preserving and disseminating valuable herbal knowledge. The initial facet involves the meticulous process of collecting herbal data, an undertaking aimed at curating and preserving the intricate details of traditional medicinal practices. Subsequently, the focus shifts to the establishment of a robust database system and design, laying the foundation for a structured repository that ensures the seamless storage and accessibility of this wealth of herbal information. Finally, the classification of the gathered information emerges as a pivotal element, orchestrating a systematic categorization that unveils the diverse medicinal properties inherent in the local herbs.

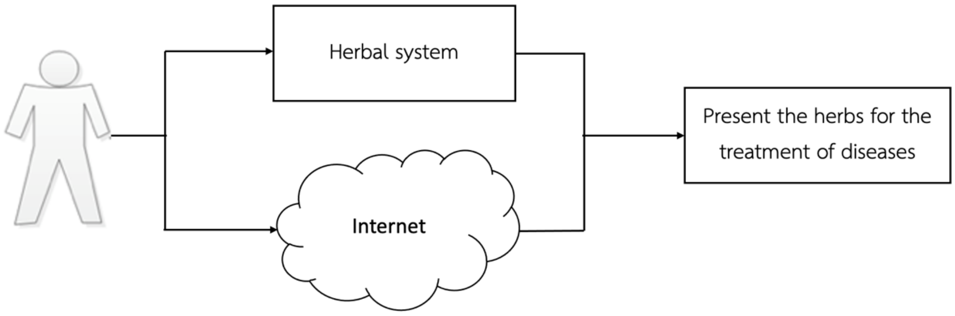

Figure 1 Piyamit Herb System Overview

System Architecture

The overall system architecture of Piyamit Herb encompasses three primary components, illustrated in Figure 1: the user interface (UI), the herbal system, and the detailed presentation of each herb. The UI serves the dual purpose of receiving information and presenting it comprehensively. As the direct point of interaction between users and the system, the UI is meticulously designed to impart a user-friendly experience. Recognizing the diverse range of devices employed by users, including personal computers, smartphones, and tablets, the UI is crafted to seamlessly accommodate all screen sizes. The challenge lies in arranging interface elements that can adapt to various layouts in real-world scenarios.

The second section is the herbal system. It is dedicated to recording the name, detailed information, and images of each herb. Developed to aggregate valuable herbal information shared by knowledgeable contributors on our web application, this section categorizes data into three sets: training, validation, and test. The training set imparts foundational knowledge for herbal classification models, while the validation set serves as a benchmark for selecting the most effective model. The test set evaluates the chosen model with a reliable quantity proportionally. The dataset is divided into training, validation, and test sets in a ratio of 80:10:10. The emphasis on a larger training set ensures the prevention of overfitting before transitioning to the validation set.

The third section is presentation of herbal details. This section provides an in-depth presentation of each herb, showcasing their medicinal properties for treating various diseases. Users can explore the rich knowledge of local herbs in Piyamit, representing a heritage passed down through generations.

Collecting Raw Data

Piyamit Herb is a vital initiative driven by the imperative to preserve the wealth of knowledge regarding local herbs in the Piyamit community, situated in Betong, Yala. The community, comprised of descendants of Malaya communists, boasts experts proficient in utilizing herbs for medicinal purposes, many of whom are currently at an average age of 80 years old. The urgency to develop Piyamit Herb stems from the realization that this invaluable knowledge, if not documented or passed on, faces the risk of being lost. Therefore, a crucial aspect of this development involves collecting knowledge from these elderly experts to ensure its transmission to the younger generation keenly interested in herbal practices.

Furthermore, the Piyamit community has evolved into a significant attraction in Betong, making Piyamit Herb not only a resource for the local populace but also a valuable trove of information for tourists and other Thai individuals. The collected data is stored in the system database accessible through the Internet, subsequently organized into a learning dataset.

The data collection process commenced with interviews involving 10 esteemed local experts from the Piyamit community, well-versed in the diverse array of herbs. These experts, formerly associated with the Malaya communists, were tasked with sharing their knowledge, with a primary focus on the types of herbs and their treatments or medicinal properties. The information gathered during interviews, initially documented by hand, underwent digital transformation, resulting in 365 digital documents.

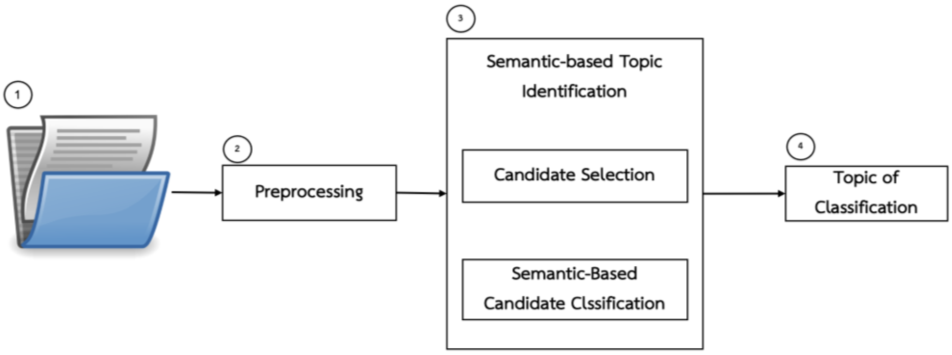

Figure 2 The overview of how to develop the local herbal data collection of Piyamit

Illustrated in Figure 2, the data processing unfolds in four key steps. The first step involves converting information from local experts into a digital format, encompassing detailed descriptions and medicinal properties of the herbs. The second step necessitates preparing the descriptive data for subsequent processing, involving essential preprocessing steps such as word segmentation, executed using the PyThaiNLP library, specifically Deepcut. After word segmentation, irrelevant words or stop words are eliminated through PyThaiNLP. The processed data yields a set of words or phrases, subsequently grouped semantically using corpora from PyThaiNLP and Wordnet. The final two steps involve identifying semantic-based topics through candidate selection and semantic-based candidate classification, focusing on grouping data based on diseases the herbs can treat. The ultimate step is to present the relationships among local herbs based on their similar properties.

Preprocessing Data

The intricacies of Thai language, particularly in conveying the detailed medicinal properties of local herbs, necessitate a meticulous data preparation process before proceeding with further analysis. Thai, being more complex than Western languages like English and other Eastern languages like Chinese and Japanese, presents unique challenges. Notably, the absence of sentence boundaries and gender indicators adds complexity to the language. Furthermore, the limited repositories of words or phrases with similar meanings in Thai pose challenges to effective word and sentence segmentation, rendering Thai language processing less effective than its English counterpart. Figure 3 outlines the critical data preparation steps preceding classification.

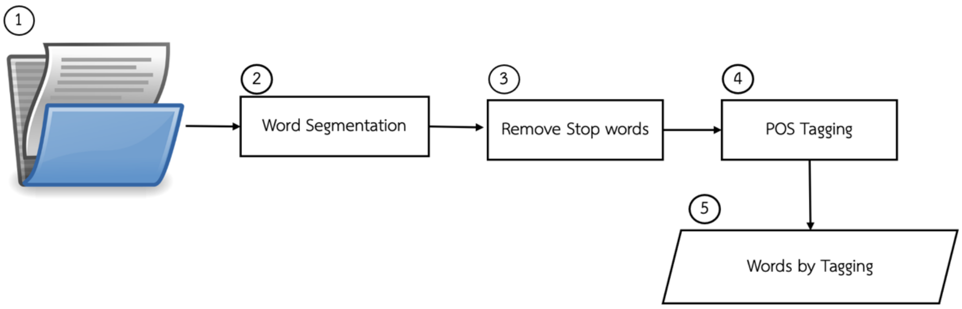

Figure 3 How to preprocess data of the local herbs of Piyamit

The preprocessing unfolds in four key steps to yield results in the form of words or phrases labeled with word characteristics or Part of Speech (POS). Initially, information gleaned from interviews with 10 local elderly experts is stored in a digital format, resulting in 365 documents covering 40 local herbs. The next step involves processing all documents with word segmentation, utilizing Deepcut—a component of PythaiNLP. Deepcut employs Convolution Neural Network (CNN) for character prediction, enhancing binary classification. Subsequently, the dataset undergoes stop word removal in the third step, aimed at excluding insignificant words that frequently appear in sentences or documents, as well as symbols used in sentences. PythaiNLP.corpus, along with imported Thai stop words, facilitates this removal. The final step encompasses Part of Speech (POS) tagging, a pivotal aspect of natural language processing. POS tagging reveals sentence patterns, elucidating the functions of words within sentences, such as suffixes, pronouns, verbs, and adverbs. Named-entity recognition is employed to recognize herb names, marking them as entities in the broader context of people, locations, organizations, products, medical codes, time expressions, quantities, monetary values, and percentages.

Semantic-Based Topic Identification

Semantic-based topic identification plays a pivotal role in the realm of natural language processing, serving as a crucial component of this research. This process involves classifying words or phrases with similar meanings, and it unfolds in two primary stages. The first stage, keyword selection, focuses on extracting candidate keywords from each document, with a specific emphasis on two types: names and adjectives. Names are chosen based on their functional significance or length, while adjectives are selected in accordance with their order of appearance in a document. Following the candidate keyword selection, the similarity of words and phrases is rigorously tested.

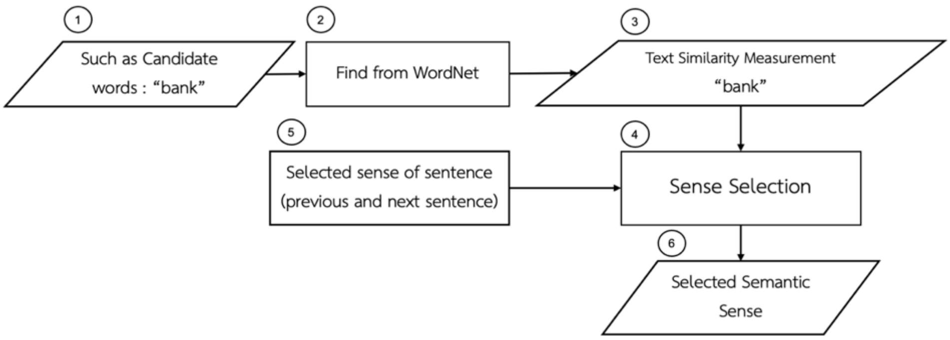

Figure 4 The steps of semantic-based topic identification

The semantic-based topic identification process in this research is structured into four key steps, as illustrated in Figure 4. Initially, candidate keywords undergo preprocessing before being selected. Subsequently, these candidate keywords are compared with words from Wordnet through a comprehensive search and verification process. In the third step, the measurement of semantic similarity or distance at the unit level is conducted. In the Thai language, this involves utilizing words or phrases from Thai Wordnet, a resource derived from Princeton’s Wordnet (PWN) Semantic Similarity. The final step, step 4, entails selecting a word with a context similar to both the preceding and subsequent sets of candidate keywords. This step is crucial due to the recurrence of selected words within a document.

Topic of Classification

Within the corpus of 356 documents detailing local herbs from Piyamit, a taxonomy of 10 distinct classes has been established for the purpose of classification. The selection of these classes is informed by the diverse properties exhibited by each herb. The defined classes are as follows:

Class 1: Blood pressure and lowering cholesterol
Class 2: Lowering sugar level in blood
Class 3: Cancers

Class 4: Rheumatoid arthritis
Class 5: Aches
Class 6: Relaxation
Class 7: Skincare
Class 8: Brain nourishment
Class 9: Male enhancement
Class 10: Female enhancement.

This classification schema provides a structured framework for organizing and understanding the diverse medicinal properties inherent in the local herbs documented within the Piyamit dataset. Each class serves as a categorical lens through which the distinct attributes and benefits of the herbs can be systematically analyzed and comprehended.

In conclusion, the tripartite composition of Piyamit Herb encapsulates a holistic approach to safeguarding and leveraging the rich tapestry of herbal wisdom within the Piyamit community. The integration of data collection, database design, and information classification culminates in a comprehensive system poised to empower both the local community and broader audiences. This multifaceted initiative not only preserves the legacy of traditional herbal knowledge but also facilitates its seamless transmission and utilization in the contemporary landscape.

Research Result

To assess the efficacy and user satisfaction of the Piyamit Herb system, we conducted evaluations involving both local youths and experts. Additionally, we measured the accuracy and precision of herb classification using specific metrics.

User Satisfaction Evaluation

A group of 20 local youths, aged between 18 and 23 years in Betong, Yala, tested the system for accuracy and design. The user satisfaction response from this group was overwhelmingly positive, with 95% expressing satisfaction with the overall system. Furthermore, a panel of 5 experts in web application development evaluated the application, resulting in a satisfaction rate of 90%.

System Performance Evaluation

To measure the accuracy and precision of herb classification, we compared the classifications generated by our model to those provided by 5 expert scientists. Three key metrics were utilized: precision, recall, and F-measure.

Precision

Precision gauges the system's ability to handle irrelevant data. Calculated as the ratio of correct predictions to the total number of predictions as shown in Equation (1).

$$precision = \frac{correct}{output-length} \qquad (1)$$

$$precision = 0.61$$

Our model achieved a precision of 0.61, indicating a 61% correctness rate.

Recall

Recall assesses the system's ability to classify relevant documents. Recall is a ratio of the amount of data retrieved to all documents as shown in Equation (2).

$$recall = \frac{correct}{reference-length} \qquad (2)$$

$$recall = 0.82$$

With a recall of 0.82, our model demonstrated effectiveness in retrieving relevant data, as the recall exceeded 0.5, signifying a good performance.

F-measure

The F-measure provides an overall evaluation by considering precision and recall, calculated as shown in Equation (3).

$$F\_measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

$$F\_measure = 0.70$$

Our model achieved an F-measure of 0.70, indicating an effectiveness of 70% in correctly classifying herbs based on their treatment.

In summary, the Piyamit Herb system received positive user satisfaction feedback and demonstrated a commendable performance in herb classification, achieving accuracy rates that affirm its reliability and utility.

Conclusion and Discussion

Piyamit Herb represents a significant accomplishment in the field of herbal information systems, contributing to knowledge preservation, local economic promotion, and broader educational outreach. Our efforts in developing the system align with the challenges identified in the literature review, particularly those associated with the complexities of the Thai language.

The system's architecture comprises three main components: a user-friendly interface catering to various devices, a robust herbal system recording detailed herb information, and a presentation section offering in-depth insights into local herbs in Piyamit. Overcoming challenges related to the Thai language, our system is designed to formalize and store herbal knowledge that was traditionally passed down verbally, ensuring accurate transmission to future generations.

Our system's success aligns with the literature's identification of challenges in Thai language processing, such as the absence of sentence boundaries and limited repositories for similar meaning words. While leveraging existing Thai word corpora from PythaiNLP and Wordnet, our work hints at the necessity for dedicated corpora tailored to the nuances of the Thai language. This implicit integration emphasizes the real-world application of insights gained from the literature.

The system evaluation involved 20 local youths, 5 web application development experts, and 5 scientists assessing herbal classification correctness. User satisfaction, gauged through the local youth group and expert evaluations, showed 95% and 90% satisfaction rates, respectively. Precision, recall, and F-measure metrics were employed to evaluate the accuracy of herb classification. Our model demonstrated a precision of 0.61, recall of 0.82, and an F-measure of 0.70, indicating the effectiveness of our system.

Piyamit Herb not only achieves its immediate goals of knowledge preservation and dissemination but also implicitly addresses the challenges outlined in the literature. The successes and insights gained from this research set the stage for ongoing improvements and future exploration in the field of herbal information systems, emphasizing the need for continuous adaptation to language nuances and the potential development of dedicated corpora for enhanced natural language processing in Thai.

As we look forward, there are key areas for future development to fortify the system. Handling increased traffic, especially with multimedia uploads, stands as a priority. The system's adaptability to slang and local language nuances also demands attention, acknowledging the diverse ways users might input information. Presently, with approximately a hundred concurrent users, Piyamit Herb operates smoothly, as affirmed by positive user experiences and satisfaction during evaluations. Beyond the local community, the system serves as a valuable resource for anyone keen on exploring traditional herbal knowledge. Piyamit Herb is not merely a local repository but a dynamic platform fostering cultural exchange. It not only caters to the needs of the present community but transcends geographical boundaries, making local wisdom accessible to a broader audience. The successful realization of this project signifies a commitment to safeguarding cultural heritage through innovative technological solutions.

In conclusion, Piyamit Herb reflects a successful synergy of tradition and technology, ensuring the enduring legacy of local herbal knowledge. As we continue to enhance the system's capabilities, we embark on a journey to empower communities and share the wealth of cultural insights embedded in traditional practices.

References

Ahmad, R., Riaz, M., Khan, A., Aljamea, A., Algheryafi, M., Sewaket, D., & Alqathama, A. (2021). Ganoderma lucidum (Reishi) an edible mushroom; a comprehensive and critical review of its nutritional, cosmeceutical, mycochemical, pharmacological, clinical, and toxicological properties. *Phytotherapy Research, 35*(11), 6030-6062.

Anaya, M., Tibaduiza, D., & Pozo, F. (2017). Detection and classification of structural changes using artificial immune systems and fuzzy clustering. *International Journal of Bio-Inspired Computation, 9*(1), 35-52.

Bharti, S., Babu, K., & Jena, S. (2017). *Automatic Keyword Extraction for Text Summarization: A Survey*. Retrieved from https://arxiv.org/abs/1704.03242.

Jang, B., Kim, M., Harerimana, G., Kang, S., & Kim, J. (2020). Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. *Applied Sciences, 10*(17), 5841.

Jurisch, M., Böhm, S., & James-Schulz, T. (2020). Applying Machine Learning for Automatic User Story Categorization in Mobile Enterprises Application Development. *International Journal of Interactive Mobile Technologies, 14*(15), 81-94.

Kurnia, R., Tangkuman, Y., & Girsang, A. (2020). Classification of User Comment Using Word2vec and SVM Classifier. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(1), 643-648.

Liang, C., Tian, D., Liu, Y., Li, H., Zhu, J., Li, M., Xin, M., & Xia, J. (2019). Review of the molecular mechanisms of Ganoderma lucidum triterpenoids: Ganoderic acids A, C2, D, F, DM, X and Y. *European Journal of Medicinal Chemistry, 174*, 130-141.

Nasar, Z., Jaffry, S., & Malik, M. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management, 56*(6), 102088.

Nokkaew, K., & Kongkachandra, R. (2018). *Keyphrase extraction as topic identification using term frequency and synonymous term grouping*. A paper presented at the 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing, Pattaya, Thailand.

Saelim, K., & Kongkachandra, R. (2017). *Semantics-based topic identication for keyphrase extraction*. A paper presented at the 12th International Conference on Knowledge, Information and Creativity Support Systems, Nagoya, Japan.

Saelim, K., Luangwiriya, T., & Kongkachandra, R. (2019). *Using Conceptual Graph to Represent Semantic Relation of Thai Facebook Posts in Marketing*. A paper presented at the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing, Chiang Mai, Thailand.

Sharma, A., Chaurasia, S., & Srivastava, D. (2020). Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec. *Procedia Computer Science, 167*, 1139-1147.

Siddiqi, S., & Sharan, A. (2015). Keyword and keyphrase extraction techniques: a literature review. *International Journal of Computer Applications in Technology, 109*(2), 18-23.

Tapsai, C., Meesad, P., & Unger, H. (2019). An Overview on the development of Thai natural language processing. *Information Technology Journal, 15*(2), 45-52.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.