International Journal of Cyber and IT Service Management (IJCITSM) p-ISSN: 2797-1325
Vol. 3 No. 2 October 2023 e-ISSN: 2808-554X

The Development of Semantic Dictionary Prototype for the Balinese Language

Cokorda Pramartha[1], I Made Yoga Mahendra[2], Gede Primahadi Wijaya Rajeg[3], I Wayan Arka[4]
Computer Science Department[1,2], English Department[3],
School of Culture, History & Language[4]
Udayana University[1,2], University of Oxford[3], Australian National University[4]
Street Unud Raya Campus, Jimbaran, Kec. South Kuta, Badung Regency, Bali 80361[1,2]
Broad St, Oxford OX1 3AZ[3], Canberra ACT 2601[4]
Indonesia[1,2], United Kingdom[3], Australia[4]
e-mail: cokorda@unud.ac.id[1*], yoga.mahendra2000@gmail.com[2],
gede.rajeg@ling-phil.ox.ac.uk[3], wayan.arka@anu.edu.au[4]

Author Notification
8 August 2023
Final Revised
25 August 2023
Published
31 August 2023

To cite this document:
Pramartha, C., Mahendra, I. M. Y., Rajeg, G. P. W. ., & Arka, I. W. The Development of Semantic Dictionary Prototype for the Balinese Language. International Journal of Cyber and IT Service Management. Retrieved from https://iiast.iaic-publisher.org/ijcitsm/index.php/IJCITSM/article/view/132

DOI:
https://doi.org/10.34306/ijcitsm.v3i2.130

Abstract

The Balinese language is the mother tongue that is used by the people who live in Bali for daily communication and social interaction. Currently, the Balinese language just like many other mother tongue languages in Indonesia is in the process of being abandoned by the younger generation for many reasons for instance it is because there are limited resources available on the Internet. Semantic ontology has been chosen as the backbone of this system due to minimising the ambiguity of the translation using the keyword-based query. The development utilises the prototyping system development approach by focusing on the involvement of the user at every stage of development. The system has three main features namely semantic browsing, semantic searching, and antonym searching. We carried out two types of evaluation (functional and non-functional) by involving 30 participants and the results show users perceive the usefulness and ease of use to be positive.

Keywords*: Ontology, Semantic, Prototyping, Balinese Language, Digital Humanities*

1. Introduction

Bali is one of the thousands of islands in Indonesia which is known for its tangible (temple, building, statue, etc) and intangible culture (cultural practices). In Bali, the Balinese language has been utilised for communication purposes either in formal or informal settings. Indonesia also recognises as a country with a diverse ethnic group (more than 600 ethnic groups) and has around 719 mother tongue languages. To unite all ethnic group, this country use the national language (Bahasa Indonesia) as an official language at school and government office. With the spread use of the national language across the country has affected the use of the mother tongue language in the community [1, 2]. The Balinese language just like many other mother tongue languages in Indonesia is in the process of being abandoned by the younger generation because its use is narrow (mainly at home). Also, because of the availability of language resources that can be used to learn about this language available on the Internet. To be overwhelmed by this issue it is necessary to develop a semantic system [3] that can be utilised by the younger generation to preserve and continue the use of the Balinese language. Therefore, it is necessary to build a knowledge system for the Balinese language [4].

In preserving and continuing the use of this language by the younger generation, the available information and learning material in the digital is certainly needed. For this kind of system, the Semantic ontology has been chosen as the backbone of this system due to minimising the ambiguity of the translation using the keyword-based query [5]. In this study, the authors developed an ontology that would then be applied to the Balinese Language Search system which is divided into three layers of the Balinese language (high tongue, common tongue, and low tongue).

2. Research Methods

This research is using the Design Science Research Methodology (DSRM) as the research method. The DSRM method offers a useful approach to conducting research aimed at creating and evaluating information technology (IT) artefacts to address problems. As shown in Figure 1, DSRM has several stages, namely (1) Problem identification and motivation; (2) Objectives for solutions; (3) Design and development; (4) Demonstration and Evaluation; and (5) Communication [6, 7].

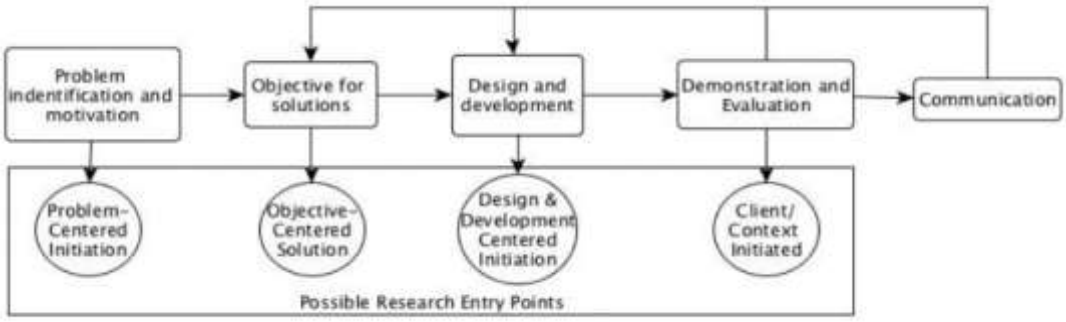

Figure 1. Stages in Design Science Research Methodology

2.1. Problem Identification and Motivation

The first stage of the DSRM is the stage for identifying the problems raised in this research. The basic problem in this study is the lack of available online resources about Balinese language knowledge that can be utilised by the younger generation to learn and continue the use of the Balinese language. As previously mentioned, the Balinese language consists of three different layers (high tongue, common tongue, and low tongue), the synonym words are completely different among these layers [8].

2.2. Objective for Solution

This stage is the stage of determining solutions that are used to overcome problems in the previous stage (identification and motivation). A possible solution is by creating an application that can help users to search and understand the words in the Balinese language. The system to be built uses ontology as a backbone because ontology can be used to present information in a more semantic (meaningful) and it can be mapped into a structured and systematic collection of information sources [9].

2.3. Design and Development

The prototyping system development approach has been utilised to develop the Information Technology (IT) artifact [10] that consists of five stages (Requirement analysis, System design, Develop prototype, Evaluation and improvement, and Implementation) as shown in Figure 2. The advantage of this method is that feedback is obtained rapidly from users because of their involvement in every step of the system development [11].

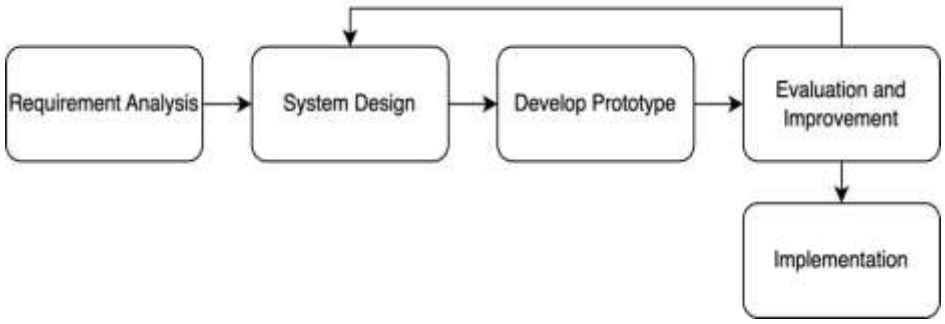

Figure 2. Stages in Prototyping Development Method

2.3.1. Requirement Analysis

Requirements analysis in systems engineering and software engineering focuses on the tasks that determine the needs or conditions to meet the new or altered product or project, considering the potentially conflicting requirements of the various stakeholders, analysing, documenting, validating, and managing software or system requirements. The study of requirements is crucial to the success or failure of a systems or software project. The requirements must be written, actionable, quantifiable, tested, traceable, relevant to identified business needs or opportunities, and detailed enough for the system design [12]. For this study, we analyze the form of functional requirements and non-functional requirements.

2.3.2. System Design

The system is designed to be worked on the top of the Internet, which means can be accessed anytime and anywhere using any Internet browser. The cloud-computing services have been utilised to implement the design. Moreover, the user interface development of the system features is also carried out during this stage as shown in Figure 3. Two main features of browsing and searching are designed to work in the proposed system.

Figure 3. User Interface Design for Three Main Features (browsing, searching, and antonym)

The antonym feature development is using the Levenshtein Distance algorithm as a method for correcting the spelling input by the user. First, by calculating the distance between the input words and the words available in the ontology and looking at the two strings a (input word by user) and b (word available in the ontology) with the following formula:

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad \text{.....(1)}$$

Second, the calculation output that has a distance = 0 mean that the word is successfully found. However, if the distance number is >= 1 & <= 2 the result of the word will be displayed as a suggestion for the word entered by the user, whereas if the distance is > 2 then the suggestion will not be displayed.

2.3.3. Develop Prototype

Once the requirement analysis and system design have been completed, we continue to develop the prototype in the form of a mid-fidelity prototype, so that users can interact with the system. The involvement of the user will be very crucial in this step, they will help the developer to improve the prototype to meet the requirement.

2.4. System Evaluation

The prototyping evaluation stage aims to ensure that the developed prototype meets the requirement. For this purpose, we conduct two main evaluations: 1) The Black-box testing is to evaluate the functionality of the system, and 2) The Technology Acceptance Model (TAM) approach is to evaluate the non-functional requirement of the system [13, 14].

2.5. Communication

The final stage in the development of this system is to document all information and knowledge related to research so that the results of the research can be stored in written form that can be published in scientific journals or any scientific outlet.

3. Results and Discussion

3.1. System Design and Development

As mentioned in the previous section, the Design Science Research Methodology method requires the design and development of the proposed system which will be discussed in the next section.

3.1.1. Data

The data collection carried out in this study was knowledge data related to the level of language in Balinese (high tongue, common tongue, and low tongue). The data was obtained by conducting observations and interviews with Balinese language experts who are recognised the experts in this domain. Moreover, text resources were taken from the available Balinese language book and hardcopy of the Balinese dictionary. After all the data has been collected, the data is stored in a spreadsheet and then populated to the ontology. Data collection is also carried out by looking for important points fundamental to constructing ontologies. There are 35 words of data have been used in developing the basic model of our ontology that is taken from the Swadesh–Yakhontov list that can be seen in Figure 4.

1. Saya	2. Anda	3. Ini
4. Siapa	5. Apa	6. Satu
7. Dua	8. Ikan	9. Anjing
10. Kutu	11. Darah	12. Tulang
13. Telur	14. Tanduk	15. Ekor
16. Telinga	17. Mata	18. Hidung
19. Gigi	20. Lidah	21. Tangan
22. Tahu	23. Mati	24. Beri
25. Matahari	26. Bulan	27. Air
28. Garam	29. Batu	30. Angin
31. Api	32. Tahun	33. Penuh
34. Baru	35. Nama	

Figure 4. Swadesh-Yakhontov List (Indonesia language)

3.1.2. Ontology Model Development

The Semantic ontology has been chosen as the backbone of this system due to minimising the ambiguity of the translation using the keyword-based query [5]. The Menthontology ontology development method is used to construct our semantic ontology for the Balinese language. This methodology is a classic method to develop ontology models and has the advantage of providing details of every process that is carried out. In addition, the methontology also provides an ability to reuse any available ontologies for further development [3, 15]. In this section, we describe the stages of Menthontology to develop our ontology:

a) Specification

In this stage, a description of the Balinese language ontology is produced, namely the Balinese language domain, purpose to develop an ontology model that makes it easy to classify words in Balinese, the scope of research on words in Balinese and sources of knowledge used through the available literature and serial of interviews.

b) Knowledge Acquisition

At this point, the techniques that we use to acquire Balinese ontology knowledge are discussions and interviews with Balinese language experts. The interview and discussion mainly used the Balinese language. Moreover, we transcribe the outcomes of the interview and conduct text analysis. This analysis aims to study the flow of concepts explained during the interview and from the book/literature that we studied. Also, performing text analysis to provide an understanding of the structures of the language that can be mapped in the ontology as the definitions, assertions, concepts, attributes, values, and relationships.

c) Conceptualization

After the knowledge acquisition is completed, we move to formalise it into a triplet (Subject-Predicate-Object) in the ontology. For instance the word “to give” (*ngamaang* in Balinese) as a subject, is-a (predicate), verb (object) as shown in Figure 5.

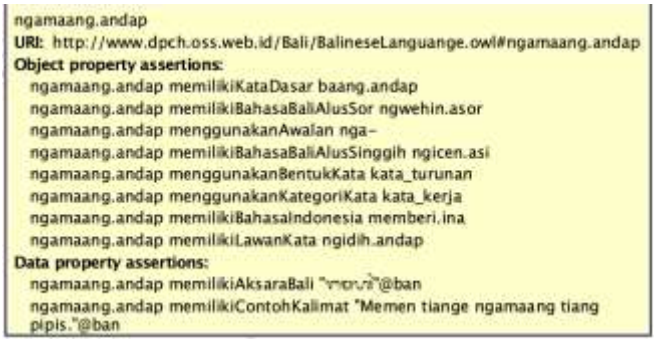

Figure 5. Conceptualization

d) Integration

In this stage, we integrate the ontology model created with the available ontology that has been created by another researcher [5] and then discussed further with the ontology engineering and Balinese language experts.

e) Implementation

The ontology conceptual design has been completed using the Methontology method and then implemented using Protégé 5.5.0 ontology editor. In the Protégé 5.5.0 software, each part of the ontology is given a meaning according to the results of the task flow in the Methodology, concepts are defined as classes, ad-hoc binary relations are defined as object properties, and instances are defined as individuals. Figure 6 shows the visualization result of the Balinese language ontology in the form of an ontograph. The ontograph in the Balinese language ontology defines 21 object properties, 222 individuals, 1 data property and 20 classes.

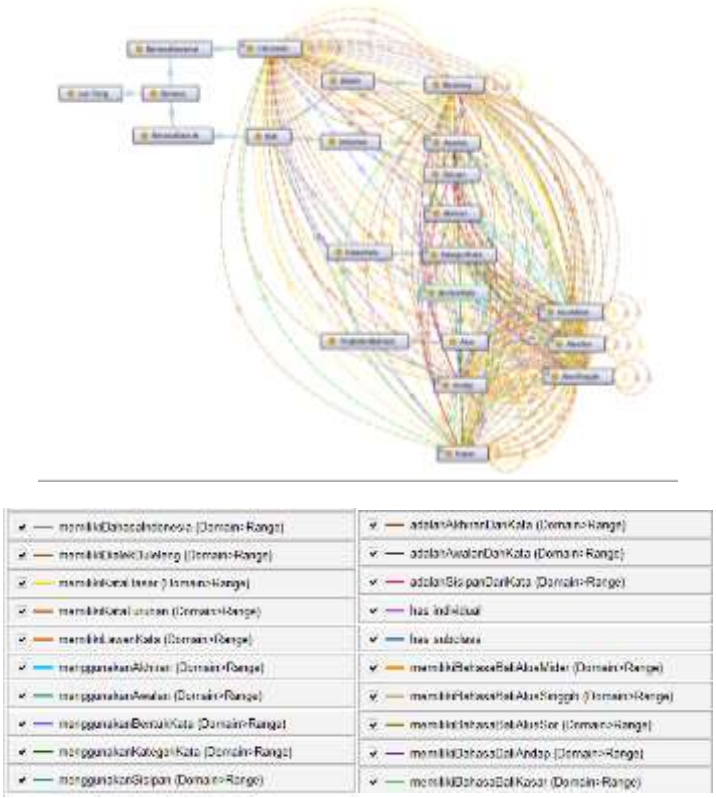

Figure 6. Ontographs in the Balinese Language Ontology

f) Evaluation Stage

We evaluate the ontology formal model that has passed the development stage using the Pellet Reasoner tool to check the consistency of the ontology. From the reasoning process carried out, the ontology of Balinese words has been consistent (see Figure 7)

```
INFO  16:19:21 ------------------------------ Running Reasoner ------------------------------
INFO  16:19:22 Pre-computing inferences:
INFO  16:19:22     - class hierarchy
INFO  16:19:22     - object property hierarchy
INFO  16:19:22     - data property hierarchy
INFO  16:19:22     - class assertions
INFO  16:19:22     - object property assertions
INFO  16:19:22     - same individuals
INFO  16:19:23 Ontologies processed in 2024 ms by Pellet
INFO  16:19:23
INFO  17:36:15 ------------------------------ Running Reasoner ------------------------------
INFO  17:36:15 Pre-computing inferences:
INFO  17:36:15     - class hierarchy
INFO  17:36:15     - object property hierarchy
INFO  17:36:15     - data property hierarchy
INFO  17:36:15     - class assertions
INFO  17:36:15     - object property assertions
INFO  17:36:15     - same individuals
INFO  17:36:15 Ontologies processed in 74 ms by Pellet
INFO  17:36:15
```

Figure 7. Balinese Language Ontology Reasoning Output

3.2. System Implementation

In the implementation stage, the system is divided into two: first, the process of connecting the ontology to the system, and second, implementing the browsing, searching, and antonym word features. The current prototype is available online at https://kamusbali.oss.web.id that can be accessed using any Internet browser.

3.2.1. Ontology Implementation

In developing and storing an Ontology-based website, a triplestore server is needed to manage the data in the Ontology. The Apache Jena Fuseki is chosen as a server to manage the Ontology. Once the server has been installed over the cloud computing infrastructure, then the created ontology is sent to this triple store as shown in Figure 8.

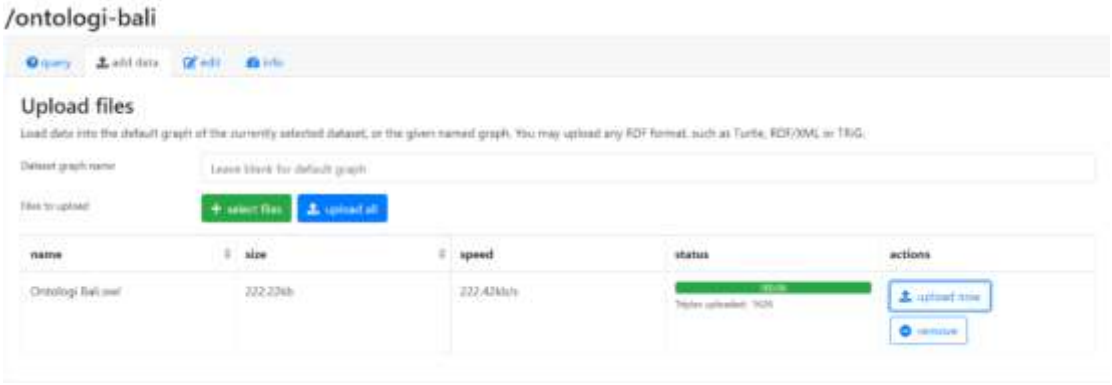

Figure 8. Uploading Ontology to the Triplestore Server

3.2.2. Interface Implementation

The user interface is important for our system because it provides direct contact to the user. Figure 9 shows the landing page and browsing interface of the prototype. On the home menu provide a brief description of the prototype. On the browsing page, the user can browse the information based on the desired criteria. The criteria referred to the level of language, word forms and word categories.

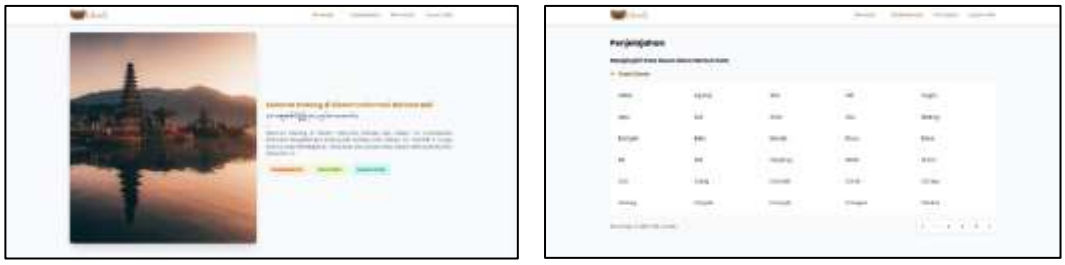

Figure 9. Landing Page and Browsing Interface

Figure 10 shows the semantic searching and text-based searching interface. On the semantic search page feature, users can search for Balinese knowledge information by specifying the desired word criteria based on the structure of the ontology. While on text-based searching, the user provides any text in the text fields. Once the request uses semantic search or text-based search by the user then the query will match the ontology modelling using the selected keywords and then display to the user.

Figure 10. Semantic Searching And Text Based Searching Interface

The antonym Balinese word feature interface (see Figure 11 a), there is an input to find antonym Balinese words. On the word detail page, if the user selects a word in browsing, the search or Antonym word will be directed to the detail page. Figure 11 b shows the detail of every word in Balinese.

Figure 11. (a) Antonym and (b) Word Details Interface

3.3. System Evaluation

Evaluation performs to check whether the developed system is meet the requirements. First, we conducted a functional evaluation using the Black-box evaluation. Black-Box testing aims to carry out functional testing of all the features in the system. The features being evaluated are semantic browsing, semantic searching, and antonym feature. We focussed only on the input data and output display to the user. Second, we conducted non-functional requirement using the Technology Acceptance Model (TAM) approach by providing task and questionnaire to the users.

3.3.1. Functional Evaluation

As previously mentioned, our prototype provides three features (searching, browsing, and antonym) that can be utilised by the user. The Black-box method approach has been adopted that focuses on the input and the output of the system. To evaluate each of the features we created a test code as shown in Table 1, and we expect all of those features to work well before launch to the public. To use those features, the user does not need to register with the system.

Table 1. Black Box Evaluation Code

Test Name	Test Code	User	Expected Results
Searching	P1	*Guest User*	Succeed
Browsing	P2	*Guest User*	Succeed
Antonym	P3	*Guest User*	Succeed

Table 2 shows the results of all three main features (searching, browsing, and antonym) evaluation. We focus on four aspects of the feature functionality such as a landing page, input, output, and display of the page when it accesses through different platforms (mobile, tablet, and PC).

Table 2. Semantic Searching Evaluation

No	Scenario	Expected results	Results	Conclusion
1	Landing page	The system can display the landing page for the three features	True	Succeed
2	Input function	The input function able to be used for the three features	True	Succeed
3	Output function	The system can display output based on user input	True	Succeed
4	Display function	The display of the system is responsive, accessible on mobile, tablets and PCs	True	Succeed

3.3.2. Non-Functional Evaluation

In the information systems domain, the Technology Acceptance Model (TAM) is a classic model that is commonly adopted to evaluate user experience in the new IT [16, 17]. To explain users' attitudes toward using an information system, this model uses perceived usefulness and perceived ease of use as the variables [18]. According to the model, when users perceive the system as useful and easy to use, then their attitude toward using the system will be positive. For this evaluation, we invited participants throughout the Udayana University network and 30 participants responded to our invitation. All participants were briefed about the purpose of their involvement and what they would do in this research.

First, participants were asked to perform five browsing tasks (explore the World Wide Web by following one interesting link to another) on the browsing function. In each navigation task, the participant must answer a question by constructing two elements of the query using the proposed prototype of the navigation function.

Second, the participants had to perform five search tasks (requesting an item of information from the database) search function on the proposed system. To be able to answer questions using the search function, a participant must construct two or more query components as input filters and generate category queries from the ontology hierarchy by doing filter the output and tracking by clicking the search button.

Third, similar to the browsing and searching tasks, participants were required to perform five tasks in finding the answer to the antonym query. They need to write a word that has been provided by the researchers and find the answer using the proposed system.

After the participants finished performing the browsing, searching, and antonym tasks, we marked each one of the submissions. We classified the marking skim into three categories:

1. Incorrect. This score is awarded when the participant does not give the correct answer to the assigned task.
2. Partially correct. This score is assigned when the answer meets one of the required criteria.
3. Correct. This point is assigned if the participant gives a completely correct answer to the question.

Table 3 shows most of the participants were able to use the system and find the correct answer with more than 80% correct on each of the task categories.

Table 3. The output of the system after performing the task.

Category	Searching	Browsing	Antonym
Incorrect	6.67%	0.67%	8%
Partly correct	9.33%	9.33%	0%
Correct	84%	90%	92%

After performing the browsing and searching tasks, all participants were invited to answer a small set of questions regarding the ease of use and usefulness of the digital portal. We adopted the questionnaire constructed by Davis [19], in which we focussed on two dimensions: perceived usefulness (PU) and perceived ease of use (PE). The items are measured using 7-point Likert scale (strongly agree = 7, agree = 6, somewhat agree = 5, neither agree nor disagree = 4, somewhat disagree = 3, disagree =2, and strongly disagree =1).

After completing the navigation and search tasks, all participants were asked to answer a series of short questions regarding the ease of use and usefulness of the digital portal. We used a questionnaire developed by Davis [19], in which we focused on two aspects: perceived usefulness (PU) and perceived ease of use (PE). Items were measured using a 7-point Likert scale (strongly agree = 7, agree = 6, slightly agree = 5, neither agree nor disagree = 4, slightly disagree = 3, disagree = 2 and strongly disagree = 1).

The results indicate that the usefulness (PU) and ease of use (PE) perceived by the users are statistically significantly higher than the average score of 4.0 (see Figure 12.). In addition, with more than 97% of the total users responding toward agree scale, the results show

that in general, users find the usefulness and ease of use of the semantic dictionary prototype to be positive.

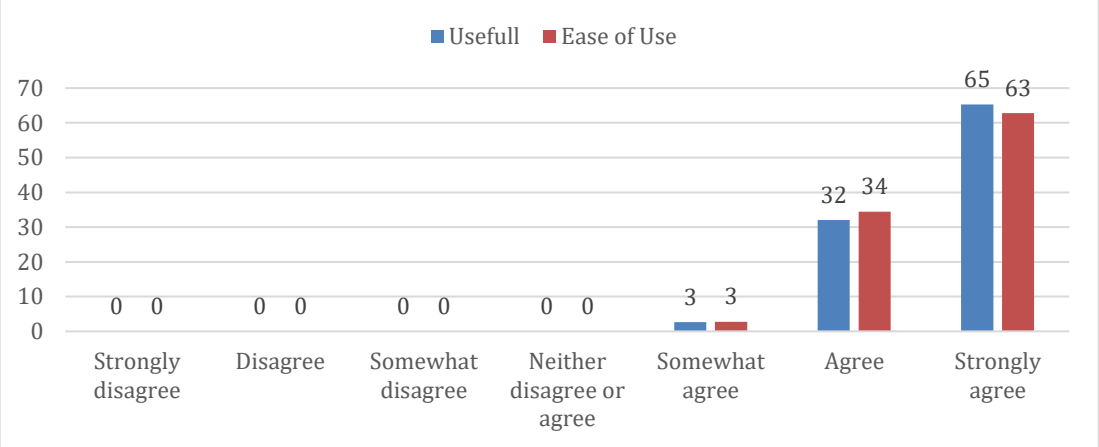

Figure 12. Perceived Usefulness and Perceived Ease of Use

4. Conclusion

We have shown our research on one aspect of Bali's cultural heritage, the Balinese language. Our contributions include the development, implementation and evaluation of a semantic dictionary prototype for the Balinese language. The semantic ontology has been chosen as the backbone of this system due to minimising the ambiguity of the translation using the keyword-based query. Our study involved 30 participants from the Udayana University network to evaluate the key features of our prototype system "semantic browse, semantic search, and antonym functions" in terms of ease of use and usefulness. The results show that, in general, users perceive the usefulness and ease of use of the semantic dictionary prototype for the Balinese language to be positive.

Acknowledgement

This project received funding from the 2022 Udayana University International Research Collaboration Grant [grant no. B/78.191/UN14.4.A/PT.01.03/2022]

Reference

[1] N. L. S. Beratha, N. W. Sukarini, and I. M. Rajeg, "Balinese Language Ecology: Study about language diversity in tourism area at Ubud village," *Jurnal Kajian Bali,* vol. 7, no. 2, pp. 121-134, 02 October 2017.

[2] N. N. A. Suciartini, "Eksistensi Bahasa Bali Di Ranah Milenial (Studi Kasus Kemunculan Parodi Hai Puja)," *Kamaya: Jurnal Ilmu Agama,* vol. 1, no. 2, pp. 134-150, 2018.

[3] C. Pramartha, I. Koten, I. G. N. A. C. Putra, I. W. Supriana, and I. W. Arka, "Pengembangan Sistem Dokumentasi Melalui Pendekatan Ontologi untuk Praktek Budaya Bali," *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI,* vol. 11, no. 3, pp. 259-268, 12/27 2022.

[4] C. Pramartha, I. B. A. I. Iswara, and I. K. A. Mogi, "Digital Humanities: Community Participation in the Balinese Language Digital Dictionary," *Jurnal Sistem Informasi (Journal of Information System),* vol. 16, no. 2, pp. 18-30, 10/31 2020.

[5] C. Pramartha, I. W. Arka, K. K. Kuan, and I. Darmawan, "Developing Semantic Ontology for Practical Digital Balinese Dictionary," in *Proceeding International Conference on Information Technology, Multimedia, Architecture, Design, and E-Business*, 2022, vol. 2, pp. 16-22.

[6] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of management information systems,* vol. 24, no. 3, pp. 45-77, 2007.
[7] C. Pramartha, J. G. Davis, and K. K. Y. Kuan, "Digital Preservation of Cultural Heritage: An Ontology-Based Approach," in *The 28th Australasian Conference on Information Systems*, Hobart, Australia, 2017, pp. 1-12, 2017.
[8] M. Covarrubias, *Island of Bali*. Singapore: Periplus Editions (HK) Limited, 2008.
[9] A. Nguyen, L. Gardner, and D. Sheridan, "Towards ontology-based design science research for knowledge accumulation and evolution," 2019.
[10] C. Pramartha and N. P. S. H. Mimba, "Udayana University International Student Management: A Business Process Reengineering Approach," *ComTech: Computer, Mathematics and Engineering Applications,* vol. 11, no. 2, pp. 57-64, 2020.
[11] P. López, J. Mabe, L. Etxeberria, and E. Gorritxategi, "Iterative Prototyping Methodology for the Development of Innovative and Dependable Complex Embedded Systems Through SPC&KPI Techniques," Cham, 2018, pp. 65-80: Springer International Publishing.
[12] M. Kraeling and L. Tania, "2 - Software Development Process," in *Software Engineering for Embedded Systems (Second Edition)*, R. Oshana and M. Kraeling, Eds.: Newnes, 2019, pp. 33-87.
[13] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly,* vol. 13, no. 3, pp. 319-340, 1989.
[14] I. M. W. Wirawan and C. Pramartha, "Pengembangan Sistem Informasi Penanganan Penderita Gangguan Jiwa Dengan Pendekatan Enteprise Systems," *SINTECH (Science and Information Technology) Journal,* vol. 5, no. 1, pp. 31-41, 04/21 2022.
[15] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," presented at the AAAI-97 Spring Symposium Series, Stanford University, EEUU, 1997. Available: http://oa.upm.es/5484/
[16] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: a comparison of two theoretical models," *Management science,* vol. 35, no. 8, pp. 982-1003, 1989.
[17] R. L. Hasanah, F. F. Wati, and D. Riana, "TAM Analysis on The Factors Affecting Admission of Students for Ruangguru Application," *Jurnal Sistem Informasi (Journal of Information System),* vol. 15, no. 2, pp. 1-14, 10/31 2019.
[18] I. M. Suarta and I. K. Suwintana, "THE INFLUENCE OF E-LEARNING CHARACTERISTICS AND BASIC ICT COMPETENCIES TO ACTUAL USAGE OF E-LEARNING: A PATH DIAGRAM MODEL," *Jurnal Sistem Informasi (Journal of Information System),* vol. 11, no. 2, pp. 59 - 67, 10/29 2015.
[19] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly,* pp. 319-340, 1989.