Melayu Grid: Sharing Language Service to Bridge Cultural Difference

Arif Bramantoro
Faculty of Computing and Information Technology in Rabigh
King Abdulaziz University
Jeddah, Saudi Arabia
asoegihad@kau.edu.sa

Irfan Syamsuddin
Department of Computer and Networking Engineering
Politeknik Negeri Ujung Pandang
Makassar, Indonesia
irfans@poliupg.ac.id

***Abstract*—Today grid computing powered by web service technology has been a research interest for decades. Several grids of language services are available in Internet as part of daily activity. Melayu Grid is proposed to share Indonesian, Malay and regional language resources. To ease the sharing, each language is located as a distinctive concept in language service ontology and eventually measured its semantic closeness to other languages. By utilizing semantic closeness, it is expected that the endangered language resource that has less amount of its speaker can take advantage by replacing it with other language resource that has better resources and speakers. This project implements service grid architecture as a culture grid for Indonesia and Malaysia to balance cultural differences and conflicts in these two neighboring countries. Multicriteria optimization is utilized to adapt the multi culture feature of several language services. It substitutes several regional languages services that are coming from the same language families or having similar linguistic vocabularies, syntax, morphology and semantics.**

***Index Terms*—Grid Computing, Multicriteria Optimization, Semantic Closeness, Intercultural Collaboration**

I. Introduction

Recently, there has been a growing trend to combine grid computing with web service technology in Internet community. A grid of language services is one of the example. It provides a convenient tool for daily activities in a borderless society. It provides a motivation to have a grid of language services specific to particular countries.

Some countries have regional languages besides their national and commonly used immigrant language. Providing the right language services for users with different culture is the key of sharing language resources. It aims at increasing the usability of language services by connecting them to different service domains. This paper proposes to have specific culture grid, Melayu Grid, for accommodating culture service domain for two neighboring countries, Indonesia and Malaysia. Melayu Grid is important to improve the relationship between the two countries. Melayu Grid is also important to reduce the extinction possibility of the endangered languages.

In Melayu Grid, similar words having different meaning and connotation are wrapped as warning services. Claimed heritages represented in cultural terms are wrapped as sensitive services. Claimed cultural contents are wrapped as audio, image and video services. These services are shared with language services previously developed in [1] to improve the quality of communication between users in the two countries.

The contribution of this paper is three-fold. First, the replacement of unavailable shared language services is enabled through the use of semantic relatedness based on language family ontology. Second, an architecture of service grid for Melayu Grid is provided so that it can connect other service grids. Third, the composition of services in the Melayu Grid is optimized by utilizing multicriteria optimization.

II. Cultural difference

The conflict between Indonesia and Malaysia spans territorial problems, culture claims, and even artist marriages. For example, when there is Malaysian who said ”Pakai telpon saya, bisa percuma berbual 30 minit” which means ”Use my phone, it can be 30 minutes free talk” to Indonesian counterpart, it may cause a problem. In Indonesian, ”percuma berbual 30 minit” means ”big talk useless in 30 minutes.”

Communication problem is not just about the translation problem. It can be caused by societal function. For example, there is a discussion between several Javanese persons and one Malaysian. Javanese persons use Indonesian language and sometimes English to their Malaysian counterpart. However, Javanese persons tend to use Javanese language amongst themselves. This is due to the Javanese culture that considers as not polite if not using Javanese regional language when they meet another Javanese. It triggers the culture difference when a Javanese asks his Javanese friends ”Sesuk gelem belonjo?” which means ”Wanna shopping tomorrow?” Malaysian who does not understand Javanese language may guess from the word ”belonjo” which means ”shopping”. The word ”belonjo” sounds like ”belanja” in Indonesian and can be translated as ”treating” in Malaysian. It may make a Malaysian think his Javanese counterpart as not fair since he treats only Javanese friend. Since it is a Malaysian culture to not expose his objection, it turns out to be not a smooth conversation.

On the other hand, several people speak more than one language in some regions. It is common in some countries that there are additional languages; known as regional language, instead of their national language. In Indonesia, there are

978-1-7281-5467-1/20/$31.00 ©2020 IEEE

722 regional languages and it is estimated that 168 of these languages will extinct in the next 20 years [2]. As illustrated in Fig. 1, regional languages in Indonesia are not distributed evenly. From west to east, the population number decreases but the number of regional languages increases. It means that most regional languages are endangered since only few speakers left. These regional languages are different from Indonesian national language and immigrant languages such as Arabic, European and Chinese. It becomes worse when the usability of the language is decreasing by a mixed marriage or migration. People who married with others in different ethnic or migrated to different place as a tendency to reduce the use of regional languages. They also stop teaching their children the regional languages.

Melayu Grid aims at sharing language services in Malaysia and Indonesia, including regional languages. By sharing regional languages, it contributes in preserving regional languages. Melayu Grid may not increase the number of persons who use the regional language, but it can add the number of activities in which the language is used.

The focus is on Indonesian reginal languages since the population of Malaysian is mainly from Indonesian immigrants who still maintain their regional languages in Indonesia. By still maintaining their regional languages, immigrants in Malaysia is also practising the culture which they or their ancestors previously belonged to. For example, some Malaysians migrated from Java islands in Indonesia who use Javanese as their secondary language are still using Kebaya, a Javanese traditional clothes, and inherited this culture to their descendants. This evidence shows that culture is strongly related to language.

Each regional language in Indonesia can be classified to different language family. From 722 regional languages, there are 25 families as illustrated in Fig. 2. The classification of the languages into language family is based on genetic classification scheme. However, two languages in the same language family can be completely different and difficult to be used by the speakers of the two languages to communicate each other.

This research uses an ontology of language family to increase the granularity of language services. Each language family is divided into several sub-families with different levels in ontology. Languages established in a lower region of ontology has a tendency to be less distant than those established in a higher region of the same ontology. This ontology is useful to select language that is unavailable due to limited resources or speakers. Limited language resources can be resources without mother-tongue speakers or even with no known speakers. An example of language family ontology for Austronesian family is partly described in Fig. 3.

In term of web service, this ontology is used as service ontology by considering each service which has input language as its input and targeted language as its output. For example, Indonesian to Javanese dictionary service has Indonesian language a service input and Javanese language as service output. Each service input and output is located at the lowest level of language family ontology. This research modifies the service interface to accommodate not only language as input language and targeted language, but also language family. For example, service user or designer can request a dictionary service that has a Sundanese family language instead of Badui or Sunda language as its service output for Indonesian terms. By increasing the granularity of language service inputs or outputs, it is more flexible for user to choose which service appropriate for their requirements.

TABLE I
THE QUANTIFICATION OF SEMANTIC CLOSENESS

	Shared service	Replacing service	Result
1	Malayo-Sumbawan	Malayo-Sumbawan	100%
2	Malayo-Sumbawan	Malayo-Polynesian	82.5%
3	Malayo-Sumbawan	Sundanese	75%
4	Badui	Sundanese	92.5%

The same formula proposed in [4] is used to compute the semantic closeness, but the implementation is simplified into single ontology instead of multi ontology. This is due the fact that all language resources in the world are currently conceptualized in one ontology. In addition, this paper applies the semantic closeness in the actual need of sharing language services within ontology. The computation is considered useful to replace shared services. Due to the page limitation, this paper only considers the input parameters from the language family ontology. The quantification of semantic closeness between a shared service and replacing language service is shown in Table 1.

In Case 1, user shares Malayo-Sumbawan while Malayo-Sumbawan is also capable to replace. Since the same service input is belong to the shared and replacing services, the maximum score is achieved (100%). In Case 2, user shares Malayo-Sumbawan while Malayo-Polynesian is capable to replace. Since Malayo-Polynesian is the parent of Malayo-Sumbawan in the language service ontology, a high score 82.5% is obtained. A quantification of semantic closeness enables an obtrusive contrast between these two cases. In case 3, user shares Malayo-Sumbawan while Sundanese is capable to replace. Since Sundanese is the child of Malayo-Sumbawan in the language family ontology, a considerably high 75% score is achieved. The semantic closeness highlights that the relationship between Malayo-Sumbawan and Malayo-Polynesian is stronger than the one between Malayo-Sumbawan and Sundanese. This is due to the fact that Sundanese represents a distinctive meaning that can not be extrapolated to all Malayo-Sumbawan, while in Case 2 Malayo-Sumbawan is a Malayo-Polynesian and Malayo-Sumbawan has a broader meaning than Sundanese. In case 4, user shares Badui while Sundanese is capable to replace. Since Sundanese is the parent of Badui in the language family ontology, almost maximum score 92.5% is obtained. It delivers more similarity than Case 2 because the languages are established in the bottom region of the ontology.

New language services can be created by composing several language services in different workflow controls, such as

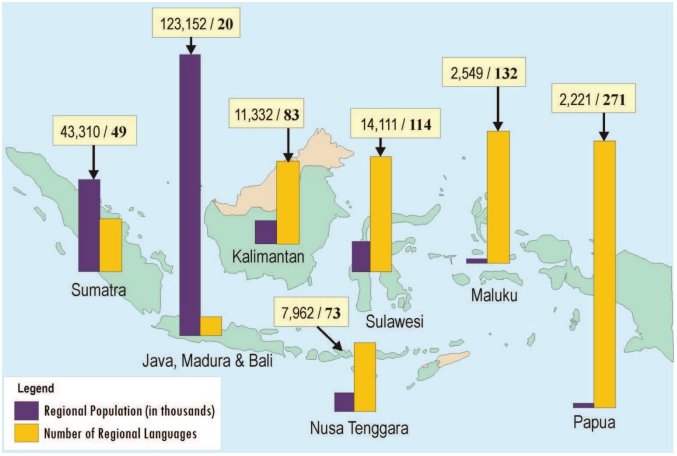

Fig. 1. A distribution of Indonesian regional languages [2]

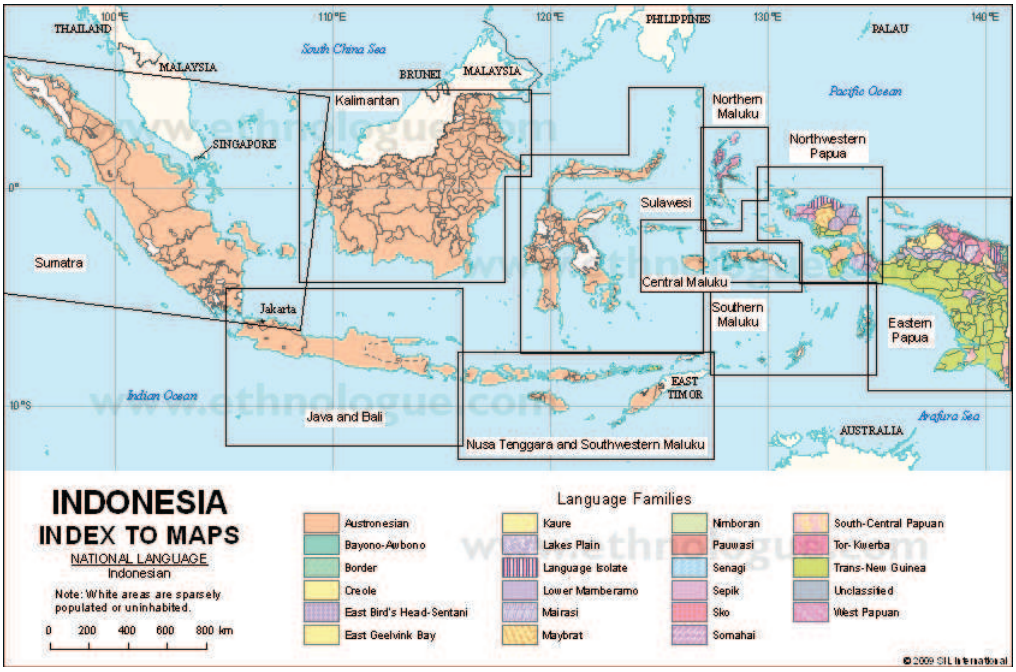

Fig. 2. A classification of Indonesian regional languages based on language families [3]

sequence, parallel and loop. For example, a new translation service is created between Malay and Javanese (one of Indonesian regional language) by composing two translation services, Malay-Indonesian and Indonesian-Javanese translation services, in one workflow. The ontology is useful for service workflow combination. There will be more new language services available in Internet by using language family as service input and output.

Some additional tasks are required to implement the use of language family ontology for service composition in this research. Each regional language is wrapped as language services conforming service interface provided by Melayu Grid and tagged with an attribute of its language family. An extended version of web service description language that supports a semantic capability creates a service ontology for language family. When the web service composition engine could not find a requested language service, it calls semantic engine to reason other nearest services located in the ontology.

III. Melayu Grid Architecture

Melayu Grid uses service grid architecture which is available as an open source architecture. Service grid is an extension of the Language Grid [5], the pioneer of service platform for language resources. The Language Grid has successfully shared more than 100 language services and 148 organizations from 18 countries could enjoy the sharing. Unlike the Language Grid, service grid is able to connect and share

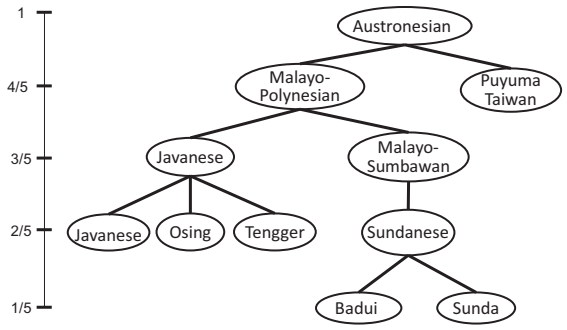

Fig. 3. Part of language family ontology for Austronesian

web services other than language services. The aim of service grid is to leverage the serviceability of language resources by allowing other service domain users access language service for different purposes.

Each service grid is federated with other service grids to optimize the use of services across domains without worrying intellectual property sharing problems. E-learning grid implements service grid to access e-learning materials via web services and connects the Language Grid to translate the materials in different languages. Medical knowledge service grid realizes service grid to improve health service by providing medical knowledge as web services and can be shared as additional medical training materials in medical grid. Agriculture knowledge service grid provides an accessible agriculture knowledge amongst farmers and uses the Language Grid to provide a knowledge transfer between farmers in different countries.

Melayu Grid is another implementation of service grid as culture grid between Indonesian and Malaysian cultures. [6] defines cultural grid as an intrapersonal grid to overcome multicultural conflicts. Here, the culture grid is considered as a grid computing in sharing culture and language services. More specifically, the idea of Melayu Grid is to show the similarities between Indonesian and Malaysian cultures through language service.

It is possible to use the Language Grid to share language services between these two cultures. However, it is difficult to accommodate language services which are sensitive to culture. For example, when a translation service is utilized in the Language Grid to translate Al-Qur'an; a holy book of the major religion in Indonesia and Malaysia, the translation of this holy book is required to explain the root culture between the two cultures. The problem is that the translation result is unsecured. The translated manuscript of holy book in particular religion is still considered as a translation, not the holy book. Although the translation accuracy is 100%, it is not the same as the holy book initially written in one specific language. Another problem is that it is not possible to control the combination of several manuscripts from different holy books in the Language Grid. With this combination, anyone can creates new religions based on the combined language services from several holy books. This responsibility that cannot be handled by the Language Grid

An architecture of Melayu Grid is illustrated in Fig. 4. The architecture consists of four main components, Service Manager, Service Invoker, Intra/Inter Grid Composer and Culture Knowledge. Service Manager is used to register and monitor culture services. Service Invoker connects user application and services. Intra/Inter Grid Composer is as a bridge with other service grids. All knowledge information about Indonesia and Malaysia culture is stored in Culture Knowledge component. This component consists of semantic engine to reason similar services based on service ontology. The culture contents supporting this component consists of culture sensitive languages and their families, common heritages and religion. Past issues considered as sensitive issues between Indonesia and Malaysia are underlying the knowledge.

Fig. 5 shows composite and atomic service available to users. Composite service engine handles the composition of regional language services. All regional language services can be composed based on their closeness relation in language family ontology. The reasoning of the closeness relation is tackled inside Melayu Grid's semantic engine. Service user does not have to know the language family knowledge since this is the domain of linguistic expert. User only has to define his requirements and tasks, composite service engine tackles the composition and querying the relationship between services to semantic engine.

Atomic service engine contains cultural text analysis service to utilize an information analysis from the composed service developed in [7]. This service is specifically applied on cultural texts and documents. Cultural warning service provides an alert to user when their language services contain cultural issues that might reduce the communication quality. Sensitive service provides several heritages information such as dancing, song, and clothes; that are claimed in both countries. Cultural video, song and image service provide a visualization of cultural contents to service users to improve the understanding of the differences within the cultural contents while using the language services.

IV. Multicriteria optimization

In natural language processing, the accuracy and completeness are two important factors in using and sharing the processing tool. If the processing tool is not accurate, no one will use it. In Indonesian regional language, low usability of the language means the potential extinction of it. Hence, the accuracy and completeness of the Indonesian regional language resources and their composition is required to avoid the language extinction. The only way to accommodate the requirement of language accuracy and completeness in web services technology is by using multicriteria optimization for quality of service (QoS) approach.

In grid computing, current QoS technique is related to network domain. QoS metrics are adopted from the network metrics, such as response time, reliability, and availability [8]. It is difficult to define language accuracy and completeness

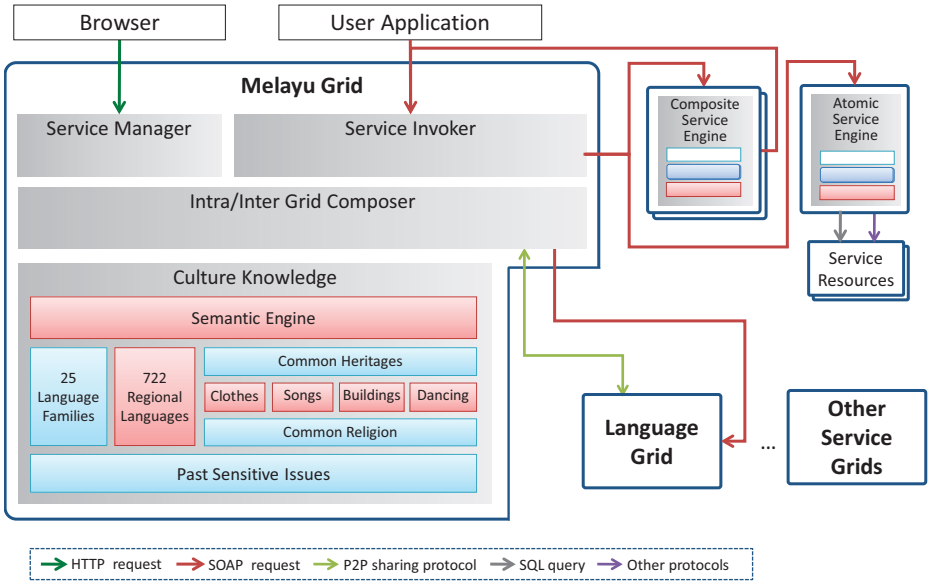

Fig. 4. An architecture of Melayu Grid

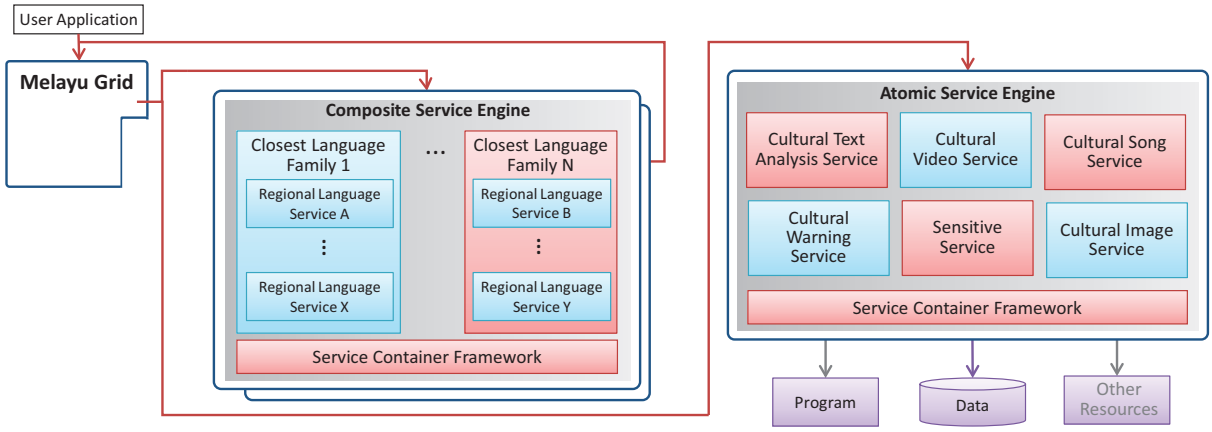

Fig. 5. Composite and atomic services architecture

metric for language services. It is also not easy to calculate QoS that accommodates language service requirement by using current QoS technique. It requires multicriteria optimization initially proposed by Pareto [9].

Fig. 6 shows an example of multicriteria problem for two different cultures. The services are represented as points in two dimensional space. The coordinate of the point corresponds to the values of the service to each culture. The services on the border shows that more than one option is available for service users. This area is called Nondominated border. Service *C* is not dominated by service *D* as well as service *E* and *F*. However, service *A* and *B* is dominated by others because there are other services that offer a better value for culture *1* or culture *2*. Sometimes several tradeoff solutions are considered more interesting to different cultures than only one extreme solution. Service user is unable to synthesize the functions into only one. This is due the characteristic of the culture that it cannot be compared, but it cannot be combined either.

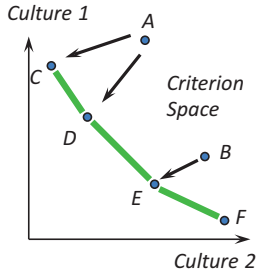

Fig. 6. Multicriteria problem for two cultures

To increase user satisfaction on culture accuracy and com-

pleteness, a multicriteria optimization is proposed for each atomic service during composition. In service composition, user skills and knowledge in particular regional, national and other immigrant languages are taken in main consideration in composing the right services. Moreover, multicriteria optimization is used to substitute several regional languages services that are descendants from the same language families or having similar linguistic vocabularies, syntax, morphology and semantics.

The formulization of multicriteria optimization is similar to the one proposed in [10]. The difference is that it focuses on Pareto optimal that accommodates culture information, which is described as follows:

- (X, D, P, C) is a group of variables to incorporate the environment of culture service
- $X = \{X_1, \ldots, X_n\}$ is a group of abstract culture services
- $D = \{D_1, \ldots, D_n\}$ where D_i a group of culture services X_i
  $D_i = \{s_{i1}, \ldots, s_{ik}\}$ where s_{ij} is a concrete culture service of the X_i
- $P = \{P_1, \ldots, P_m\}$ is a group of culture feature from each user
- $C = \{C_1, \ldots, C_p\}$ is a group of constraints which contains:
  - CS as a group of soft constraints with a penalty of $\rho C_i \in [0, 1]$
  - CH as a group of hard constraints
- $F(X) = (F_1(X), \ldots, F_q(X))$ is a utility function where $F_i(X)$ is for user culture *i*
- A solution is an accomplished determination $X = \{X_1, \ldots, X_n\}$ and solution X is Pareto optimal if and only if $\forall X', F(X) \not< F(X')$
- Nondominated border is the set $F(X)|\forall X', F(X) \not< F(X')$

For example, the formulization of two cultures between Indonesia and Malaysia is as follows:

- $X = \{X_1, X_2, X_3, X_4, X_5\}$
  - X_1: Translation service between Indonesian and Malaysian User
  - X_2: Dictionary service for Indonesian
  - X_3: Dictionary service for Indonesian
  - X_4: Multimedia service for Malaysian
  - X_5: Multimedia service for Malaysian
- $D = \{D_1, D_2, D_3, D_4, D_5\}$
  - D_1: {ms-id translation, no service}
  - D_2: {ms-id dictionary, no service}
  - D_3: {ms-en dictionary, no service}
  - D_4: {on, off}
  - D_5: {on, off}
- $P = \{P_1, P_2\}$
  - P_1: Profile of Indonesian User $P_1.religion, P_1.ethnic$
  - P_2: Profile of Malaysian User $P_2.religion, P_2.ethnic$
- $C = \{C_1, C_2, C_3, C_4\}$
  - $C_1 : X_2 =$ no service $\Rightarrow X_4$=off
  - $C_2 : X_3 =$ no service $\Rightarrow X_5$=off
  - $C_3 : P_1.religion! = P_2.religion \Rightarrow X_4 = on \wedge X_5$=on
  - $C_4 : P_1.ethnic! = P_2.ethnic \Rightarrow X_4 = on \wedge X_5$=on
- $F(X) = (F_1(X), F_2(X))$
  - $F_1(X)$: Best composition for Indonesian culture
  - $F_2(X)$: Best composition for Malaysian culture

V. Conclusion

Web service is a promising technique for sharing language resources dependent to a culture. An example of collecting and sharing language resources in two countries, Indonesia and Malaysia is provided. It is found that these language resources are sensitive to the culture where the language belonged to. These language resources may trigger a big cultural difference if they are not managed properly. A web service based system, Melayu Grid, is provided not only to share the language resources but also to contribute in reducing culture difference implied in the language resources. Melayu Grid is also useful in preserving regional languages, including the endangered ones, by increasing the usability of the languages. The usability of the regional languages increases as the languages resources are wrapped as language services and connected to other services in different service grid. If the regional language service is not available, Melayu Grid replaces the service with other service by using semantic closeness in language family ontology. Multicriteria optimization is proposed to increase user satisfaction in using Melayu Grid.

References

[1] A. Bramantoro, U. Schafer, and T. Ishida, "Towards an integrated architecture for composite language services and multiple linguistic processing components," in *International Conference on Language Resources and Evaluation*, Valetta, Malta, May 2010, pp. 3506–3511.

[2] M. Lauder, "Obstacles to creating an inventory of languages in indonesia: A dialectology perspective," *Clevedon: Multilingual Matters Ltd*, vol. 134, pp. 40–53, 2006.

[3] M. P. Lewis, *Ethnologue: Languages of the world*. SIL International Dallas, TX, 2009. [Online]. Available: http://www.ethnologue.com

[4] A. Bramantoro, R. K. Alshammari, and A. O. Almagrabi, "Measuring business entities in multi-ontologies," *International Journal of Electronic Business*, vol. 13, no. 4, pp. 342–358, 2017.

[5] T. Ishida, Y. Murakami, D. Lin, T. Nakaguchi, and M. Otani, "Language service infrastructure on the web: the language grid," *Computer*, vol. 51, no. 6, pp. 72–81, 2018.

[6] P. Pedersen, "Mediating multicultural conflict by separating behaviors from expectations in a cultural grid," *International journal of intercultural relations*, vol. 17, no. 3, pp. 343–353, 1993.

[7] Y. Murakami, M. Tanaka, A. Bramantoro, and K. Zettsu, "Data-centered service composition for information analysis," in *IEEE Ninth International Conference on Services Computing*. IEEE, 2012, pp. 602–608.

[8] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "QoS-aware middleware for web services composition," *IEEE Transactions on Software Engineering*, vol. 30, no. 5, pp. 311–327, 2004.

[9] E. Rollon and J. Larrosa, "Bucket elimination for multiobjective optimization problems," *Journal of Heuristics*, vol. 12, no. 4, pp. 307–328, 2006.

[10] R. Ramacher and L. Mönch, "Robust multi-criteria service composition in information systems," *Business & Information Systems Engineering*, vol. 6, no. 3, pp. 141–151, 2014.