Harnessing Collective Intelligence to Preserve and Learn Endangered Languages

Asfahaan Mirza(✉) and David Sundaram

Department of Information Systems and Operations Management,
University of Auckland, Auckland, New Zealand
asfahaan@mirzabros.com, d.sundaram@auckland.ac.nz

Abstract. The languages are disappearing at a frightening rate; half of 7105 plus languages spoken today may disappear by end of this century. When a language becomes extinct, communities lose their cultural identity, practices tied to a language and intellectual wealth. The rapid loss of languages motivates this study.

Current language revitalization approaches and systems are focused on creating language learning resources from already documented artifacts rather than a holistic approach towards language preservation, curation and learning. Harnessing and leveraging collective intelligence within communities through cooperation, coordination and cognition can help preserve endangered languages.

In this paper we first introduce collective intelligence, endangered languages, and language revitalization. Secondly we discuss and explore how to leverage collective intelligence to preserve, curate, discover, learn, share and eventually revitalize endangered languages. Thirdly we compare and synthesize existing language preservation and learning systems. Finally, this chapter proposes the design and implementation of "Save Lingo" and "Learn Lingo" mobile apps for preserving and revitalizing endangered languages. The systems are instantiated and validated in context of te reo Māori, which is the native language of the Māori population of New Zealand.

Keywords: Collective intelligence · Language revitalization · Endangered languages · Crowd sourced · Language learning · Mobile apps

1 Introduction

Language is a living and dynamic phenomenon that plays a significant role in our daily lives. It is essential for acquisition, accumulation, maintenance, and transmission of human knowledge regarding the natural environment and ways of interacting with it [1]. Language defines a culture, through which people communicate. Currently, languages are at greater risk of extinction than species of animals and plants [2]. Most linguists believe that at least half of the world's 7105 languages are endangered and they are not being learnt as first languages by children, and ultimately leading to the death of languages.

Firstly, to revitalizing a language, we need to document the language by capturing, curating and preserving various language artifacts such as words, phrases, songs,

P.C. Vinh et al. (Eds.): ICTCC 2016, LNICST 168, pp. 224–236, 2016.
DOI: 10.1007/978-3-319-46909-6_21

idioms, stories and dialects for future use. Secondly, the preserved language can be disseminated and learnt by the wider community. The key success to any revitalization efforts is the contribution and collaboration by its native community. Hence, leveraging collective intelligence within the community will significantly help in preserving and learning the language. In the subsequent section we explore the fundamentals of collective intelligence, endangered languages and language revitalization efforts.

1.1 Collective Intelligence

Intelligence can be defined as the ability to see things, understand and reason at different capacities. It involves one being aware and conscious of individual actions and analysis of creative and critical thinking. In this context, collective intelligence can be described as the ability for different individuals to come together and form a group with the intention to share a common line of thought [3]. Lévy (1997) defines collective intelligence as a form of subjective mobilization, highly individual as well as ethical and cooperative.

Collective intelligence can either be beneficial or negative depending on the objectives set by the group. In this digitally enabled generation, the Internet is described as a useful tool that can be used to promote collective intelligence [5]. Currently, many individuals and groups are using the Internet to collectively provide content on various issues defining global trends as well as raising awareness on topics. Search engines and social media platforms are playing a significant role in facilitating these actions.

Collective intelligence dwells on three key principles that include cooperation, coordination and cognition [6]. Using these key principles, harnessing collective intelligence enables individuals and groups to solve practical problems. In this research we explore how to leverage concepts of collective intelligence to save endangered languages. Moreover, design and implement systems that can be used to mitigate the loss of languages and revitalize the affected languages [7].

1.2 Endangered Languages

Language is a primary means of interacting among people in various forms such as in person, writing, over the phone or the Internet. We all are so called ‘social animals’ therefore communication plays a very important role in our daily living, and language enables a person to express their feelings and opinions. Krauss (1992) estimates that 90 % of the world’s 7105 languages would become endangered or extinct by the end of this century, if no language revitalization efforts are made [8]. Endangered languages are disappearing at a frightening rate; one language every two weeks [9]. The reasons for language endangerment are complex but generally it is linked to communities abandoning their minority native languages to a mainstream language that is more economically, politically and socially powerful.

Languages are one the richest part of human diversity. Currently there are 7105 living languages spoken by a world population of 6,716,664,407 [10]. The language

distribution among the world population is heavily skewed. Approximately 79.5 % of the world population speaks only 75 languages. In contrary, 3894 smaller languages are spoken by only 0.13 % of the world population [1]. The rapid decline of languages highlights the need to revitalize endangered languages for the survival of culture, diversity and knowledge.

1.3 Language Revitalization

Language revitalization is to reverse the decline of a language from becoming extinct or endangered. Linguists have proposed various models for language revitalization [11, 12]. The language revitalization models include school-based programs (total and partial immersion), children's programs outside the school, adult language programs, documentation and materials development, home-based programs, and language reclamation.

Language preservation is one of the essential tasks when revitalizing a language. Language can be documented in multiple formats including, audio and video recordings, scanned images, or written notes. The documented data can then be archived and mobilized into various publications (print and digital). The preservation is highly dependent on the younger generation learning and using their indigenous language [13]. Most of the language revitalization models focus on language learning.

Language learning is an emerging research area known as "Language Acquisition" which overlaps in linguistics and psychology. We will focus on language learning techniques that assist in language revitalization efforts. Language learning is a key component to language revitalization, as teaching the future generation their indigenous language will help keep the language alive. Learning a language is dependent on the available resources for the language. Most of the language revitalization models consist of teaching and learning a language. Past research suggests that total immersion schools and classrooms have been very successful for language revitalization [12]. However, emerging technologies provide capabilities that were not available before which gives birth to new concepts towards language revitalization. In the next section we look at how we can leverage collective intelligence for language revitalization purposes.

1.4 Leveraging Collective Intelligence for Language Revitalization

One of the most important components of existence is language; it gives people a certain sense of belonging and originality. Language connects people of the same cultural values and makes them diversified to the general world. According to Evans and Levinson (2009), linguistic is the basis for cultural preference and belonging, the two authors note that the diversity offered by languages provide a crucial element in cognitive science [14]. Community driven projects are the best examples of how collective intelligence connects their interest. It is important for companies and other corporate bodies to develop programs that help connect language to daily work practice to help preserve the endangered language. Incorporating communities' practices and

language preference in the day-to-day running of the corporations provide a sustainable step of language preservation [15].

Language provides deep insight of the culture associated with the language. Thus, gathering collective intelligence from various age groups and genders will help preserve, revitalize and pass the language to future generations. The language is rapidly evolving and like biology, only the fittest survive, similar to language, only the commonly used language will survive the change. The learning and usage of a language can be made possible by harnessing collective intelligence about a particular language and disseminating the knowledge among the community. Making the language available to wider society is made possible through the use of mobile devices and social media. This will enable the community to access their language anywhere anytime.

The diagram below illustrates the research dimensions (Fig. 1). Research in the past (inner circle) has focused predominantly on the design and implementation of traditional systems to capture and curate languages by experts in limited contexts focused on standard vocabulary and media. This research (red zone) tries to address these research problems by exploring a crowd sourced approach to harness collective intelligence using ubiquitous systems to capture, curate the linguistic diversity and richness anytime anywhere. Moreover, allow end users to discover and learn to use the endangered language.

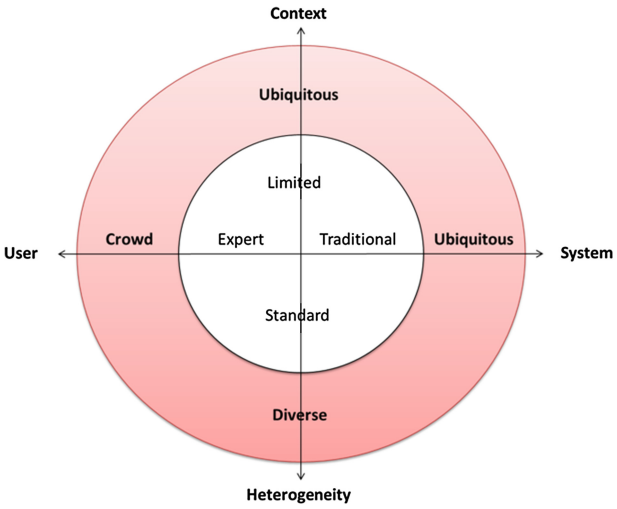

Fig. 1. Research Dimensions (Color figure online)

2 Existing Systems for Language Revitalization and Learning

Current language revitalization models and activities include school-based immersion programmes, children's programmes outside school, adult language programmes, documentation and materials development, home-based programs, and language

reclamation [1, 11, 12, 16–18]. However, the focus of these efforts have been on language learning rather than a holistic approach towards language preservation, curation, and usage. Only few indigenous communities around the world have started adopting mobile apps to revitalize their language and cultural practices [19–22]. However, the apps are mostly learning-oriented using already documented information (e.g. Go Vocab, Hika Explorer, Te reo Māori dictionary, Duo Lingo, Ojibway and Saulteaux). Ma! Iwaidja, and Duo Lingo allow data capture but does not facilitate curation process. Moreover, there is limited research on models, frameworks, and/or architectures for the design and implementation of collective intelligence based language preservation and learning system. Hence, there is a gap in existing approaches and apps, because they do not facilitate a holistic approach to preserve, curate, and learn to use an endangered language (Table 1).

Table 1. Existing systems available for revitalizing endangered languages

Functionality / **Systems**	Capture Audio	Capture Video	Capture Images	Curation Process	User Feedback	Audio Playback	Learning	Dictionary	Translator	Dialect Support	Progress Tracking	Assessments	Gamification	Leader Board	Social Media
DIXEL Mobile	N	N	N	N	N	Y	Y	Y	N	N	N	Y	Y	N	N
Duolingo	Y	N	N	N	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y
Go Vocab	N	N	N	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
Hika Explorer	N	N	N	N	N	Y	Y	N	Y	N	Y	Y	Y	N	N
iPhraseBook®	N	N	N	N	N	Y	Y	N	Y	N	Y	N	N	N	N
Language Perfect	N	N	N	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y
MindSnacks	N	N	N	N	N	Y	Y	N	N	N	Y	Y	Y	Y	N
Ma Iwaidja	Y	N	Y	N	Y	Y	Y	Y	N	N	N	N	N	N	N
Memo Cards	N	N	N	N	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y
myLanguage Pro	N	N	N	N	Y	Y	N	N	Y	Y	N	N	N	N	N
Ojibway	N	N	N	N	N	Y	Y	Y	N	N	N	N	N	N	Y
Saulteaux	Y	N	N	N	N	Y	Y	Y	N	N	N	Y	Y	N	N
Spanish 24/7	N	N	N	N	N	Y	Y	Y	N	N	Y	Y	Y	N	N
Spanish!	N	N	N	N	N	Y	Y	Y	N	N	Y	Y	Y	N	N
Te Pūmanawa	Y	N	N	N	Y	Y	Y	N	Y	N	Y	Y	Y	N	N
Touch Trainer	N	N	N	N	N	Y	Y	Y	N	N	Y	Y	Y	N	N
Tusaalanga	Y	N	N	N	N	Y	Y	Y	N	Y	Y	N	N	N	N

3 Design and Implementation of a Collective Intelligence Based Language Preservation and Learning System

Majority of languages in the world are endangered and rapidly becoming extinct. The aim of this research is to design and implement a collective intelligence driven smart mobile apps to preserve and learn endangered languages. We adopted a design science research methodology to help develop concepts, models, processes, frameworks and architectures [23–26]. Subsequently, mobile apps will be designed and implemented leveraging key principles of collective intelligence including cooperation, coordination and cognition to support vital language revitalization processes: (1) **Capture/Preserve** - words, phrases, poems, idioms, and stories as text, audio, images, and video in multiple dialects; (2) **Curate** - filter and approve captured content, and (3) **Learn and use** – context-aware dynamic games and apps based on curated data to encourage the use of the endangered language in daily life. The apps and related concepts, models, processes, and frameworks will be initially designed for preserving and learning te reo Māori, which is the native language of New Zealand. These research artefacts will then be generalised to help save and learn other dying and endangered languages.

3.1 Concepts and Processes

This research employs a holistic crowd sourced approach to harness collective intelligence to revitalize endangered languages as illustrated in Fig. 2. This model is created by synthesising concepts from collective intelligence, knowledge management, and language revitalization literature [1, 4, 5, 11, 27, 28]. It has five stages and related processes namely: capture, curate, discover, learn and share. The *capture stage* allows contributors to create/capture words, phrases, idioms, stories and songs in multiple formats including text, image and video. The *curate stage* involves moderating and refining the captured data by language experts. The *discover stage* will facilitate the wider community to retrieve knowledge from the dynamic repository. The *learn* stage allows user to learn the language using interactive games. Lastly, the *share stage* enables the dissemination of knowledge through social media among the wider community to help promote the use of language.

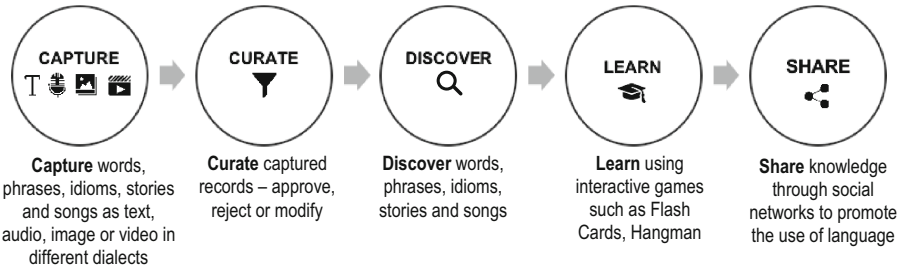

Fig. 2. Concepts and Processes to harness collective intelligence to revitalize endangered languages

3.2 Framework

The generic design elements make up the Save Lingo framework are depicted in Fig. 3. The framework incorporates collective intelligence fundamentals to save endangered languages. The framework elements support the reuse of artifacts for multiple endangered languages. Ninety percent of the artifacts such as concepts, models, processes, framework, architecture and system are not associated with the language. Only user interface changes at a system level are required to tailor it to a particular language. The artifacts are key outputs of design science research [23, 24]. The first six layers of the framework are standard and well understood in research. However, applying the fundamentals for language revitalization and learning purposes is novel.

3.3 Implementation of Language Preservation System – Save Lingo

Many indigenous communities have started to adopt Ubiquitous Information Systems and Devices (UIS&D) to preserve, maintain and revitalize their language and cultural practices [19, 20]. UIS&D refer to systems and devices (tabs, pads, or boards) that are available abundantly without boundaries [29, 30]. There is a significant rise in adoption of UIS&D among everyone; both digital natives and digital immigrants. Ubiquitous devices provide many advantages: flexibility, low cost, mobility, user-friendliness, connectivity and multimedia capabilities. These advantages significantly help in implementing Save Lingo app – holistic crowd sourced language preservation system to harness collective intelligence. The prototypical implementation of Save Lingo adapted to revitalize te reo Māori, which is the native language of the Māori population of New Zealand.

Save Lingo app as shown in Fig. 4 extends upon fundamentals from social media, knowledge management, and collective intelligence to create a highly interactive platform that allows users to remotely contribute and collaborate towards capturing, curating, discovering and sharing the endangered language. These functionalities are described in book chapter [31].

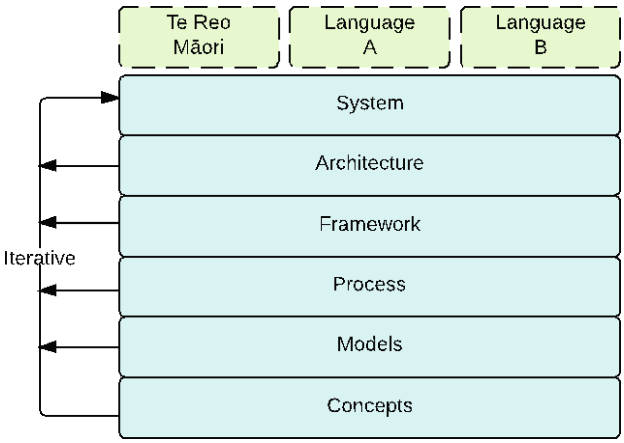

Fig. 3. Framework of generic design elements to preserve and learn endangered languages

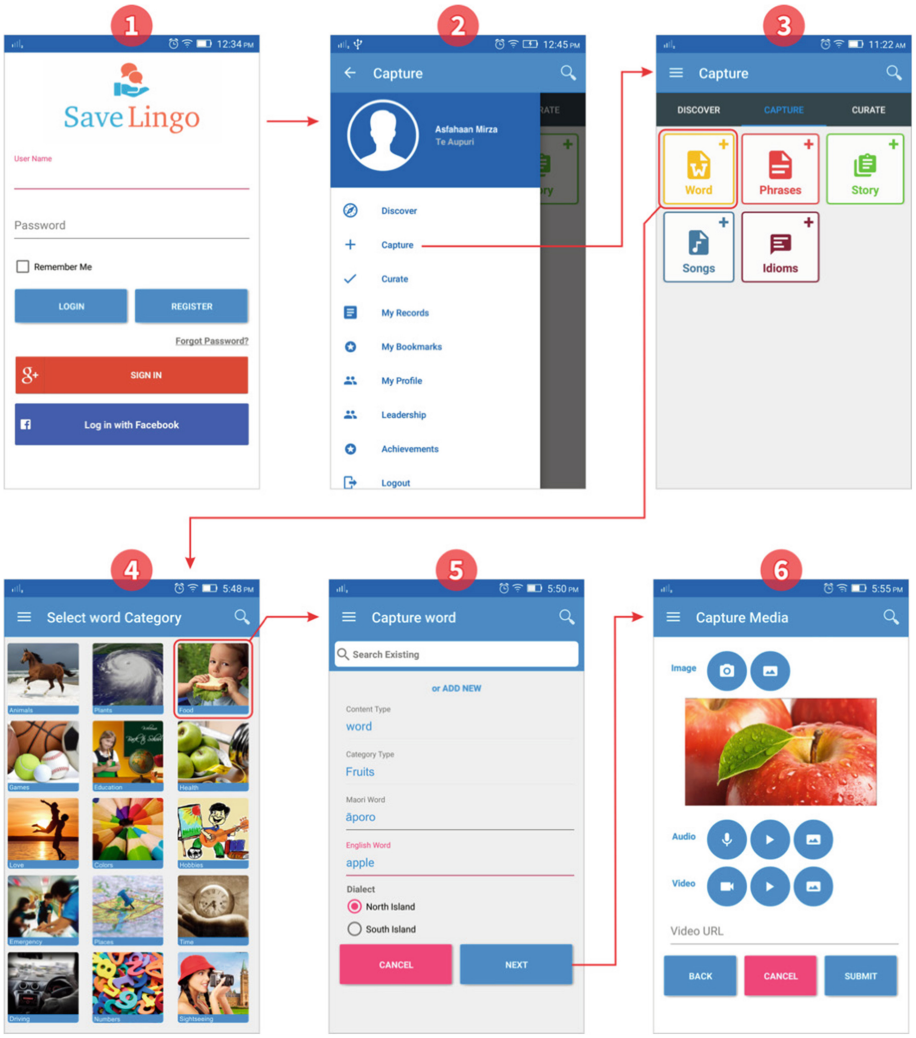

Fig. 4. Save Lingo app - Holistic crowd sourced language preservation system [31]

3.4 Implementation of Language Learning Systems – Learn Lingo

Learning a language can be a difficult task regardless of age or gender. We aim to implement Learn Lingo apps that are user-friendly, dynamic and gamified to engage, retain and help users learn an endangered language. The underlying fundamentals and learning processes, modalities and mechanisms are synthesized from language acquisition and learning literature [11, 32–36] as presented in Table 2. Every individual progresses through various learning stages to learn a new language. Each stage is unique as the modalities and mechanisms applied for learning are different. There are many learning apps available including Flash card and Hangman for various languages.

Table 2. Learning processes, modalities and mechanisms synthesized from [11, 32–36]

Stage modality	Pre-production	Early production	Speech emergence	Intermediate fluency
Listening	(Audio-Lingualism) Listen word/phrase	Listen word/phrase.	Listen phrase, story, song, idiom.	Listen phrase, story, song, idiom.
Drawing	Draw & Colour the correct answer	Draw & Colour the correct answer	Draw & Colour the correct answer	Draw & Colour the correct answer
Observe & identifying	Chose the correct image. Match the words (drag and drop)	View and select correct answer Chose the correct image. Match the words (drag and drop)	Identify and select correct phrases, story, song, idiom. Match the correct pairs (drag and drop)	Identify and select correct phrases, story, song, idiom. Match the correct pairs (drag and drop)
Reading	Repeat what's on your screen	Repeat what's on your screen	Repeat what's on your screen	Read what's on the screen and complete the comprehension activities.
Role Play	Act out the activity via Video Capture	Act out/Retell the activity via Video Capture	Act out/Retell the activity via Video Capture	Act out, Retell, Create the scenario.
Speaking	*Not Applicable*	Repeat what is on your screen Respond/Record the correct answer.	Recall, Summarize, Retell, Describe, Define, Explain, Restate. Talk to Native Speaker	Record the complete phrase, song, story, idiom. Answer or debate the question or the speaker. Group Discussion
Writing	*Not Applicable*	Complete the word by filling the missing letters for what you see or hear.	Complete the phrase/story/song by filling the missing words for what you see or hear. Define, Describe, Summarize, Explain	Translate the phrase, story, song, idiom. Analyze and correct what is displayed on the screen.
Cognitive	*Not Applicable*	List, Categories, Group, and/or Compare using a combination of speaking, listening, reading and writing	List, Categories, Group, and/or Compare using a combination of speaking, listening, reading and writing	List, Categories, Group, and/or Compare using a combination of speaking, listening, reading and writing

Nevertheless, the novel contribution of Save Lingo and Learn Lingo is that we are using the collective intelligence/data of the community which was captured and curated via Save Lingo app is used as the content for Learn Lingo's Flash Card and Hangman apps.

Flash cards to support: Observe, Identify, Listening and Speaking

Flash cards are a relatively simple way of learning key words and phrases of a language. The flash cards app supports observe, identify, listening and speaking modalities to help users learn the language. The prototypical implementation of Flash cards app is shown below in Fig. 5. The app allows user to securely login and create various flash

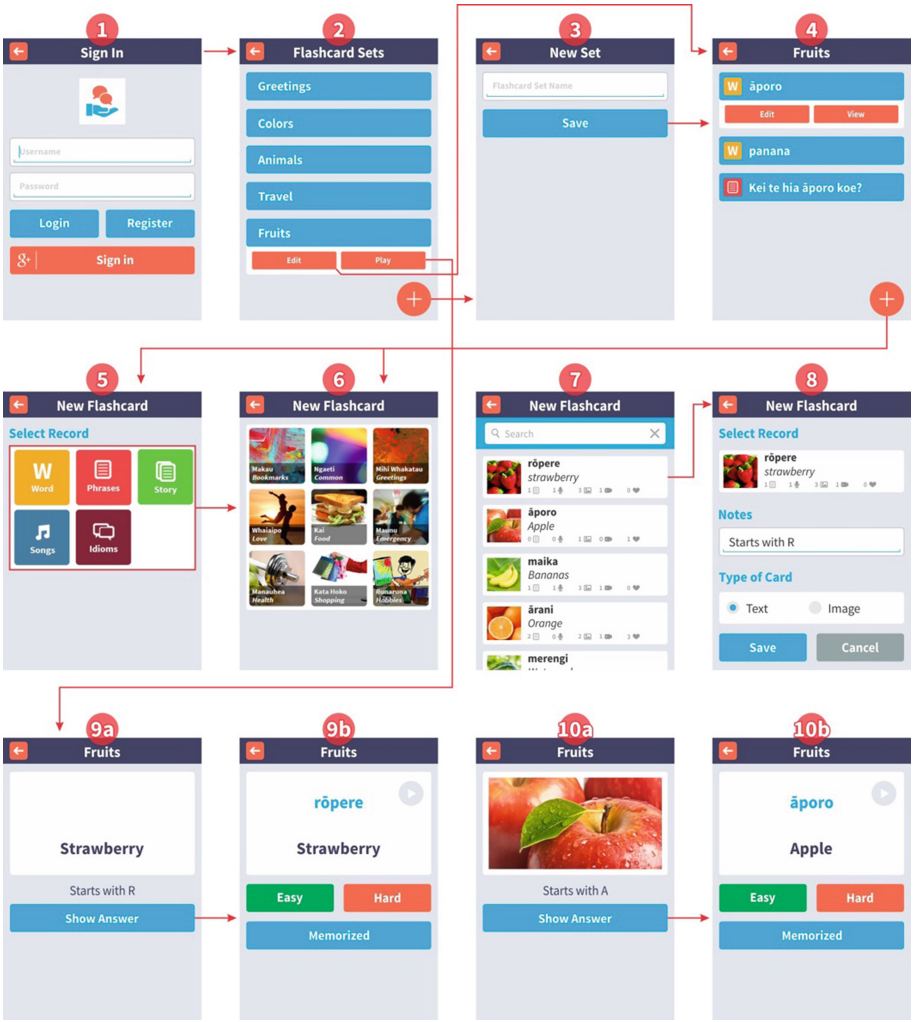

Fig. 5. Learn Lingo – Flash card app workflow

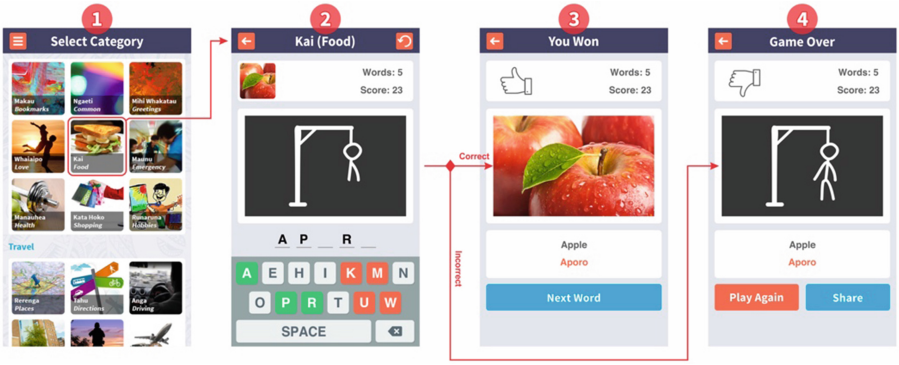

Fig. 6. Learn Lingo – Hangman app workflow

card sets. The set can be constructed using the data that was captured and curated using Save Lingo.

Hangman to support: Identify, Writing, and Reading
The Hangman app supports identify, writing and reading modalities to help users learn new words of the language. The prototypical implementation of Hangman app is displayed below in Fig. 6. The app allows user to securely login and simply select a category. The app will randomly select a curated word from the category selected by the user.

4 Conclusion

In conclusion, this research leverages from strengths of collective intelligence to address weakness in systems that attempt to preserve and/or teach languages. We first began by looking at the strengths of collective intelligence, practical problems of endangered languages and language revitalization efforts. Moreover, we explored how collective intelligence can be leveraged for language revitalization purposes. Consequently, we reviewed variety of language preservation and learning systems. Based on these finding and our previous work on Save Lingo [31], we proposed concepts, models, processes and framework to help preserve and learn endangered languages. Furthermore, we describe the implementation of Save Lingo and Learn Lingo – Flash card and Hangman, which are based on the proposed concepts, models, processes and framework in the context of an endangered language, te reo Maori. The system's functionality can easily be ported to facilitate cooperation, coordination and cognition within the community to revitalization other endangered languages. This crowd-sourced approach will harness collective intelligence to create a central repository for the distribution and revitalization of indigenous languages, knowledge, values and culture.

References

1. Romaine, S.: Preserving endangered languages. Lang. Linguist. Compass **1**, 115–132 (2007)
2. Sutherland, W.J.: Parallel extinction risk and global distribution of languages and species. Nature **423**, 276–279 (2003)
3. Kubátová, J.: Growth of collective intelligence by linking knowledge workers through social media. Lex Sci. Int. J. **1**, 135–145 (2012)
4. Lévy, P.: Collective Intelligence. Plenum/Harper Collins, New York (1997)
5. Malone, T.W., Laubacher, R., Dellarocas, C.: Harnessing crowds: Mapping the genome of collective intelligence (2009)
6. Engel, D., Woolley, A.W., Jing, L.X., Chabris, C.F., Malone, T.W.: Reading the mind in the eyes or reading between the lines? theory of mind predicts collective intelligence equally well online and face-to-face. PLoS ONE **9**, e115212 (2014)
7. Yun, J., Li, X., Wang, C.: Simulation of conducting early-warning to the endangered state of language. J. Multimedia **8**, 475–480 (2013)
8. Krauss, M.: The world's languages in crisis. Lang. (Baltim) **68**, 4–10 (1992)
9. Crystal, D.: Language Death. Cambridge University Press, Cambridge (2002)
10. Lewis, M.P., Simons, G.F., Fennig, C.D. (eds).: Ethnologue: Languages of the World. http://www.ethnologue.com
11. Hinton, L., Hale, K.L.: The Green Book of Language Revitalization in Practice. Academic Press, San Diego (2001)
12. Grenoble, L.A., Whaley, L.J.: Saving Languages: An Introduction to Language Revitalization. Cambridge University Press, Cambridge (2005)
13. Yang, M.-C., Rau, D.V.: An integrated framework for archiving, processing and developing learning materials for an endangered aboriginal language in Taiwan. ALR-2005, 14 October 2005
14. Evans, N., Levinson, S.C.: The myth of language universals: language diversity and its importance for cognitive science. Behav. Brain Sci. **32**, 429–448 (2009)
15. Lesser, E., Ranson, D., Shah, R., Pulver, B.: Collective intelligence: capitalizing on the crowd. IBM Inst. Bus. Value (2012)
16. UNESCO, Brenzinger, M., Yamamoto, A., Aikawa, N., Koundiouba, D., Minasyan, A., Dwyer, A., Grinevald, C., Krauss, M., Miyaoka, O., Sakiyama, O.: Language vitality and endangerment. UNESCO Intangible Cultural Unit - Safeguarding Endangered Languages (2003)
17. Tsunoda, T.: Language endangerment and language revitalization: an introduction. Walter de Gruyter, Berlin (2006)
18. Dwyer, A.M.: Tools and techniques for endangered-language assessment and revitalization (2011)
19. Holton, G.: The role of information technology in supporting minority and endangered languages. In: Austin, P., Sallabank, J. (eds.) The Cambridge Handbook of Endangered Languages, pp. 371–399. Cambridge University Press, Cambridge (2011)
20. Galla, C.: Indigenous language revitalization and technology from traditional to contemporary domains. Indig. Lang. Revital. Encourag. Guid. Lessons Learn. Flagstaff (Arizona, Estados Unidos) North. Arizona Univ. pp. 167–182 (2009)
21. Keegan, P., Te Taka Keegan, M.L.: Online māori resources and māori initiatives for teaching and learning: current activities, successes and future directions. Mai Rev. **1**, 1–13 (2011)
22. Nichols, D.M., Witten, I.H., Keegan, T.T., Bainbridge, D., Dewsnip, M.: Digital libraries and minority languages. New Rev. Hypermedia Multimedia **11**, 139–155 (2005)

23. Nunamaker Jr., J.F., Chen, M., Purdin, T.D.M.: Systems development in information systems research. J. Manag. Inf. Syst. **7**, 89–106 (1990)
24. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**, 75–105 (2004)
25. Baskerville, R.: What design science is not. Eur. J. Inf. Syst. **17**, 441–443 (2008)
26. Nunamaker Jr., J.F., Chen, M.: Systems development in information systems research. In: 1990 Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences, pp. 631–640. IEEE (1990)
27. Nonaka, I., Takeuchi, H., Umemoto, K.: A theory of organizational knowledge creation. Int. J. Technol. Manag. **11**, 833–845 (1996)
28. Alavi, M., Leidner, D.E.: Review: knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q. **25**, 107–136 (2001)
29. Sorensen, C., Yoo, Y., Lyytinen, K., DeGross, J.: Designing Ubiquitous Information Environments: Socio-Technical Issues and Challenges: IFIP TC8 WG 8.2 International Working Conference. Springer, Heidelberg (2005)
30. Weisser, M.: The computer for the twenty-first century. Sci. Am. **265**, 94–104 (1991)
31. Mirza, A., Sundaram, D.: Design and implementation of socially driven knowledge management systems for revitalizing endangered languages. In: Remko, H., Cranefield, J., Cranefield, J. (eds.) Social Knowledge Management in Action, in the Knowledge Management and Organizational Learning. Springer, Heidelberg (2016)
32. Viberg, O., Grönlund, Å.: Mobile assisted language learning: a literature review. In: mLearn, pp. 9–16 (2012)
34. Brown, H.D.: Principles of language learning and teaching (2000)
35. MacWhinney, B.: A unified model of language acquisition. In: Handbook of Bilingualism, p. 49 (2004)
36. Godwin-Jones, R.: Emerging technologies: mobile apps for language learning. Lang. Learn. Technol. **15**, 2–11 (2011)