Print ISSN:2717-0454
Online ISSN:2783-2856
Quarterly Journal of Innovation and Entrepreneurship
Http://journalie.ir

A Model For Developing Policy Tools to Support Language Technologies in Cyber Space

* Habibollah Asghari ** Fateme Mostajer Kheirkhah

* Assistant Professor, ICT Research Institute, ACECR, Tehran, Iran. habib.asghari@ictrc.ac.ir
** Assistant Professor, ICT Research Institute, ACECR, Tehran, Iran. kheirkhah@ictrc.ac.ir

Received: 20.01.2024 Accepted: 20.08.2024

Abstract

Language technologies are one of the important components of national authority in the cyberspace, and its promotion are the main duties of governments. The importance of language technologies makes the role of government support policies in this field an undeniable necessity. These support policies are often explained and compiled with the aim of developing practical tools in various fields such as tourism, health, education or other public areas.
In this article, the innovative approaches adopted by governments to support language technologies and writing practices in cyberspace have been investigated. In this regard, the road map and documents have been published and the development methods of infrastructure tools and practical products in the field of script and language support in the countries of India, the United States of America, Wales, Spain, and the European Union have been examined and reviewed from different aspects. Finally, the achievements of this study have been used by examining the common points and the cases that have the ability to be implemented in the case of the Persian language to provide suitable solutions for developing a national strategic document in support of the Persian language.

Keywords: Natural Language Processing, Language Technologies, Persian Language Development, Policy Tools.

Corresponding Author: Habibollah Asghari- Habib.asghari@ictrc.ac.ir

Introduction

Problem statement

The language and script of each country are among the cultural assets and valuable human heritage. In recent years, with the growth of information technology and also due to the expansion of the Internet, the boundaries of the influence of language have exceeded the geographical boundaries. In addition, the influence of language from other cultures has also doubled due to the removal of cultural boundaries. Therefore, governments have paid special attention to the support of language in cyberspace.

Natural language processing refers to the science of automatic deep understanding of the most important human production (language) with the help of computers. Language technology as a comprehensive concept includes many different applications related to language. For example, machine translation, audio recognition and processing, analysis, and extraction of information from textual data are among the sub-branches of language processing.

Purpose

Despite the growing market of industries related to natural language processing and machine translation, the investment of the public sector alongside private companies can be seen in many countries in this industry, and this is done for two main reasons: national security and market stimulation. Governments, on behalf of the people, consider themselves obliged to enter effectively and directly into discussions related to national security. The market stimulation is also raised in situations where private companies do not have the necessary motivation for investment due to economic reasons. In this case, the government would oblige itself to enter this field and compensate for this lack due to its necessity and infrastructure.

In these cases, governments resort to different methods and procedures, the simplest of which is the direct presence of the government itself to carry out this set of activities. Another view is supporting private companies to enter the desired field, which can be considered in the form of compensating a part of the companies' expenses and by creating incentives such as tax exemptions, legal concessions, etc., to provide attractions for the companies. The third is a kind of investment view in which the government invests in some targeted areas in a planned manner for the sake of long-term and public interests and in line with its governance missions.

Hypothesis

According to the points mentioned in this article, an attempt is made to examine a set of models of script and language support in cyberspace in selected countries including India, the United States of America, Wales, Spain, and the European Union. For this purpose, roadmaps and documents that have been published, as well as scientific articles from related scientific and research institutions have been studied. Finally, considering the differences in the political, cultural, and economic factors, and the learned lessons from these studies, we have developed strategies to support the Persian language in the field of cyberspace.

Background

Examining the development policies of the countries shows that, in addition to paying attention to the development of technology, the development of culture using technological tools is also of particular concern to the government. language is the main component of national culture in every country. Therefore, it is necessary to protect these basic signs of culture (Kazden, 2003).

It is essential to benefit from advanced technologies in the fields related to language. Today, there is an inseparable connection between information technology and the category of language in every country. Information technology, by developing various tools and with the capabilities it has, helps to expand and perpetuate language. Also, language is a tool for producing, developing, and disseminating content in cyberspace (Diki Kidiri, 2007).

In national development programs, special attention is paid to the national language (Stiles, 1997). Various research has been conducted for the revival and promotion of language using technological tools in other countries, among which can refer to (Warshawer, 1998), (Claire, 2010) and (Ong, 2021). National languages are so important that UNESCO has entered this field and made 2019 the year of local languages.

Methodology

This research uses review research based on library studies in which research variables have been analyzed through case studies. In this way, the development policy of several selected countries (including India, the United States of America, Wales, Spain, and the European Union) in the field of strategic and national documents and programs in support of language in cyberspace has been studied and analyzed. Based on that, the information related to these programs and the policies has been extracted in each country. This information is obtained from documents published by these governments and research articles published by researchers of these countries. Then, based on the case study, appropriate policy tools have been proposed to support the Persian language in cyberspace.

Findings

Supportive model of India

The Department of Electronics and Information Technology of the Ministry of Communication and Information Technology of India, as the custodian of information technology in this country, started the Technology Development for Indian Languages (TDIL) project in 1991. Considering the cultural and linguistic diversity in India (there are 22 languages recognized in the constitution and 12 different languages in this country). The Ministry of Communications and Information Technology of India in the TDIL project developed information processing tools and techniques to increase the influence of Information technology among different strata of people and facilitate human-machine interaction without language barriers. The comprehensive program of the Ministry of Information and Communication Technology of India for the development of language technology tools includes seven main parts as follows:

- Development of machine translation systems (MAT)
- Development of multilingual information access systems
- Development of document recognition and analysis system (OCR)
- Developing an online handwriting recognition system
- Development of a text-to-speech system for Indian languages and development of a speech recognition system for Indian languages
- Development of machine translation system for Sanskrit language

Supportive model of Wales

According to the vision of the Welsh government as explained in the strategic policy of this country for the years 2012 to 2017 under the title "One living language, one language for life" and from 2017 onwards under the title Cymraeg 2050: One million Welsh Speakers. In summary, the basic points that exist in the support plans of Wales can be mentioned as follows:

- Due to the small population of Welsh speakers, the main policy of the government's support programs is in line with the development of this language in cyberspace.
- Increasing interactions with international companies and organizations in this field
- Investing in content production in the Welsh language using language technologies
- Investing in the field of education and training of experts in related fields
- Long-term planning and budget allocation

Supportive model of USA

The United States government has supported and developed machine translation systems around the world for more than four decades. Before the famous ALPAC report in 1966, the government spent more than 13 million dollars on the development of such systems. However, the report of the ALPAC committee on that year in connection with the evaluation of machine translation and computational linguistics systems in the United States in those years introduced doubts regarding the quality and development of these systems and led to a reduction in the government budget.

Supportive model of the European Union

During the seven-year policies of FP7, H2020, and CEF, the European Union has considered comprehensive programs to support various technologies, including linguistic technology and related fields. The comprehensive program FP7 is in the years 2007 to 2013 and the comprehensive program H2020 is in the years 2014 to 2020. These supports include fields related to languages such as machine translation, voice recognition, body production, data sources, etc.

Supportive model of Spain

The Spanish government has started a comprehensive program to promote language technology since 2016, and in this program, it spends 90 million euros to improve language technology. The Ministry of Communications of Spain has organized a steering committee with the cooperation of the government organizations of this country to explain it.

The main goal of the government in this program is to prevent duplication and increase synergy through the combination of existing projects to prevent the multiplicity of organizations providing similar services. The government's general goals of this program are to increase the quality and quantity of language processing infrastructural tools for Spanish and other official languages of this country, promote the linguistics industry by facilitating the transfer of knowledge from universities to industry, help the international activities of companies active in this field, and improve It has expressed the quality and capacity of public services using language processing and machine translation technologies.

One of the challenges and issues in Iran regarding the compilation of these documents is the multiplicity of policy-makers that formulate and explain strategies to support the Persian language in cyberspace. due to this, a comprehensive document should be compiled by considering all the actors in this field.

Conclusion

In this research, the model of policy tools for projects related to natural language processing in countries of different areas of the world was investigated. The set of strategies and policies for writing and language development in this article includes a review of how support is provided to projects and products related to natural language technology in the countries of India, the United States, Wales, Spain, and the European Union.

Due to some differences in the conditions of each of these countries with Iran, a whole of the strategies mentioned in their documents cannot be applied to Iran's governance policies in cyberspace, but by adapting the existing similarities, some of the extracted goals and strategies can be used.

Reference

Cazden, C. B. (2003). Sustaining indigenous languages in cyberspace. Nurturing Native Languages, 2002(1). https://files.eric.ed.gov/fulltext/ED482034.pdf

Cohen, J. (2007). The GALE project: A description and an update. Automatic Speech Recognition & Understanding. https://doi.org/10.1109/asru.2007.4430115

COMM. (2013). LT2013 - status and potential of the European language technology markets - digital single market - European Commission. Retrieved June 4, 2016, https://economie.fgov.be/en/themes/online/strategy-european-digital/connecting-europe-facility-cef

Diki-Kidiri M (2007), Securing a place for a language in cyberspace. Paris: UNESCO. Retrieved 4 June 2016, http://unesdoc.unesco.org/images/0014/001497/149786E.pdf .

European Commission. (2014). Horizon 2020 work programme for 2014-2015. European Commission Decision C (2015)2453 of 17 April 2015. Retrieved June 4, 2016, https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-security_en.pdf

European Commission. (2015). Horizon 2020 work programme for 2016-2017. European Commission Decision C (2015)6776 of 13 October 2015. Retrieved June 4, 2016, https://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-intro_en.pdf

Kumar A., Goyal V. (2010) TDIL Programme: A Government Initiative, an International Journal of Engineering Sciences, 2010. https://ijoes.vidyapublications.com/paper/JSI/1.pdf

Olive, J. P., Christianson, C., & McCary, J. (2011). Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation. http://ci.nii.ac.jp/ncid/BB05333792

Ong, T. W. S., & Said, S. B. (2021). Selective language maintenance in multilingual Malaysia. In Advances in religious and cultural studies (ARCS) book series (pp. 207–228). https://doi.org/10.4018/978-1-7998-2959-1.ch009

Pedtke, T. R. (1997). U.S. Government Support and Use of Machine Translation: Current Status. ACL Anthology. https://aclanthology.org/1997.mtsummit-plenaries.1

United Nations. (n.d.). Protecting languages, preserving cultures | United Nations. https://www.un.org/en/desa/protecting-languages-preserving-cultures-0

Rehm, G., Berger, M., Elsholz, E., Hegele, S., Kintzel, F., Marheinecke, K., Piperidis, S., Deligiannis, M., Galanis, D., Gkirtzou, K., Labropoulou, P., Bontcheva, K., Jones, D. L., Roberts, I., Hajič, J., Hamrlová, J., Kačena, L., Choukri, K., Arranz, V., . . . Klejch, O. (2020).

European Language Grid: An Overview. arXiv (Cornell University). http://arxiv.org/pdf/2003.13551.pdf

Sepln, (2015) Plan de Impulso de las Tecnologías del Lenguaje, Retrieved June 4, 2016, http://www.sepln.org/news/plan-de-impulso-de-las-tecnologias-del-lenguaje /

St. Clair, R. N. (2010, November 17). Review of Reyhner & Lockard (2009): Indigenous Language Revitalization: Encouragement, Guidance & Lessons Learned. Language Problems and Language Planning, 34(3), 267–269. https://doi.org/10.1075/lplp.34.3.05stc.

Stiles, D. B. (1997). Four Successful Indigenous Language Programs. Teaching Indigenous Languages. https://eric.ed.gov/?id=ED415079

The Government of Spain promotes language as a factor for modernization and economic development, 2022 ,

https://www.lamoncloa.gob.es/lang/en/gobierno/councilministers/Paginas/2022/20220301_council.aspx

Vikas, O. (2001). Language Technology Development in India. Ministry of Information Technology, New Delhi, India. http://www.emille.lancs.ac.uk/lesal/omvikas.pdf

Warschauer, M. (1998). Technology and Indigenous Language Revitalization: Analyzing the experience of Hawai'i. Canadian Modern Language Review-revue Canadienne Des Langues Vivantes, 55(1), 139–159. https://doi.org/10.3138/cmlr.55.1.139

Welsh language technology action plan, October 2018 ,

https://www.gov.wales/sites/default/files/publications/2018-12/welsh-language-technology-and-digital-media-action-plan.pdf

Welsh language | GOV.WALES. (2024, February 20). GOV.WALES. http://gov.wales/topics/welshlanguage/?lang=en

White, R. M. (1993). Machine Translation Technology: A Potential Key to the Information Age. Report of the FCCSET Committee on Industry and Technology, PB-93-134336, Office of Science and Technology Policy, Washington, DC (January 1993).

اصغری، حبیب اله و محتاج، سالار، «سند ارتقاء خط و زبان فارسی در فضای مجازی»، معاونت علمی و فناوری ریاست جمهوری، ستاد توسعه فناوری اطلاعات و ارتباطات، فروردین ماه ۱۳۹۵.