USING MULTILINGUAL ANALYTICS TO EXPLORE THE USAGE OF A LEARNING PORTAL IN DEVELOPING COUNTRIES

Vassilis Protonotarios[1,2*], *Giannis Stoitsis*[1,2], *Kostas Kastrantas*[1], *Salvador Sanchez-Alonso*[2]

[1]Agro-Know Technologies, Attica, Greece

[2]University of Alcala, Madrid, Spain

ABSTRACT

Learning analytics is a domain that has been constantly evolving throughout recent years due to the acknowledgement of its importance by those using intelligent data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning [1]. Learning analytics may be applied in a variety of different cases, but their role in understanding the multilingual requirements of users of learning portals is of an outstanding significance. As the adaptation of existing portals in multilingual environments is a cost- and time-consuming aspect of the development of a portal, the outcomes of learning analytics may provide the requirements on which further multilingual services of a portal will be built, ensuring their efficiency. This paper aims to identify and interpret the behavior of users from developing countries in a multilingual learning portal using the log files of the portal by applying the methodology defined in a previous work by Stoitsis et al. [2] The paper also aims to identify the aspects that should be studied by future related works by focusing on specific regions and countries that exhibit special interest for further adaptation of the portal to additional multilingual environments.

KEY WORDS

Learning analytics, multilinguality, learning portal, developing countries

I. INTRODUCTION

In the last decade, we have witnessed a significantly increased number of interactive learning environments, learning management systems (LMS), intelligent tutoring systems, e-portfolio systems, and personal learning environments (PLE) serving the various types of education. All of these education-related products produce enormous amounts of tracking data, which in most cases is stored automatically; however, its exploitation is not fully achieved [3]. Learning analytics may be considered the study of the data sets that provide information on the interaction between learners and educational software/tools and/or educational material available online. These data sets may be used for the improvement of the learning processes in the context of almost all forms of education, including formal education [4]. Learning analytics exhibit a special interest for researchers working in this area, as they may provide substantial information on the way that the educational tools and resources are used by the intended end users. The analysis of this data may provide helpful information on the strengths and weaknesses of the material studied, identify learning patterns, predict learning outcomes, suggest relevant resources, and detect error patterns or affects of learners [5]. Since a significant number of educational resources in various formats and types are constantly developed and made available by both individuals and educational organizations, learning analytics may provide an indicator of the use and quality of these resources [6]. Various initiatives like the U.S.-based Learning Registry (http://www.learningregistry.org) have identified the importance of learning analytics and have developed a methodology that allows the collection of the corresponding data and their publication for research and application purposes.

One of the most interesting aspects of learning analytics is the information it provides about language preferences and usage. Multilinguality is one of these aspects related to learning analytics that have received a lot of attention during recent years, as it enables users to find content in their native language and at the same time use an interface that is closer to their needs and knowledge. On top of that, it minimizes the barriers raised by linguistic issues, mostly affecting users who are not able to use tools and resources in languages other than their native ones. Multilinguality is taken into consideration by the developers of web portals and e-learning services providing access to learning resources, who are usually trying to make use of the existing automatic translation technologies in order to provide access to multilingual resources and tools. Due to its importance, multilinguality is currently actively supported on both national and international levels by initiatives like the MultilingualWeb project (http://www.multilingualweb.eu), which is funded by the European Commission.

Despite the fact that in several cases the multilingual e-learning services cover a variety of different languages both at the software and the course/content level, it nonetheless seems that users from countries in dire need of these services, such as those in developing countries, may need additional support in this direction. In several cases, due to lack of funding, infrastructure, and knowledge, these countries do not develop their own e-learning material, so users from developing countries are usually depending on the e-learning material provided by other sources, such as those published online, as well as on open educational resources (OER), which are usually provided by such e-learning services. However, it is not possible for e-learning services to meet the linguistic needs of users from all different countries, covering all of the users' possible languages and dialects. Among the various workarounds used for solving this issue, there is the use of a generic automatic translation service (such as Google Translate and Microsoft Bing Translator), which is integrated into the portal. Calling its API will translate the user interface and/or content into the language defined by the user; however, since these services are not domain specific, the results may be poor, thus degrading the quality of the offered e-learning services. In other cases, both the user interface and the educational material are manually translated, therefore exhibiting higher quality and added value to the services; however, this requires the investment of a significant amount of time and money, both of which appear to be scarce, especially in the last few years. An intermediate solution is the revision/enrichment by humans of the automatically provided translations, which significantly minimizes the effort required in achieving the multilingualism required by the users.

In this direction, this paper aims to study how a learning portal serves users from developing countries and to identify their multilinguality requirements. Despite the fact that there are already publications detailing the processes for the evaluation of web portals (for example, see [7]), we decided to follow a different approach and base our analysis on the analytics of the portals. For this, we carried out an analysis of the log files of the Organic.Edunet web portal, a learning portal providing access to agricultural educational resources from a network of content providers, focusing on topics like organic agriculture, agroecology, and other green topics, such as sustainable development, energy, environment, and ecology. These log files were analyzed based on a methodology followed in other related studies [8] in order to help us identify the way that the portal is used by users from different contexts, coming from developing countries.

The remainder of the paper is structured as follows: Section 2 provides background information on multilinguality aspects and definitions, as well as on the Organic.Edunet web portal. Section 3 provides information on the methodology used in this study, including the general context of the analysis, methods, and tools, as well as the variables studied. In Section 4, the results of the study are presented, and they are discussed within the context of this work in Section 5. Section 6 summarizes the conclusions of this work and points to the open issues that require further work.

II. BACKGROUND

A. The Organic.Edunet Case Study

This paper focuses on the Organic.Edunet web portal (www.organic-edunet.eu), a multilingual learning portal deployed in 2010, which provides access to almost 11,000 digital educational resources in the fields of organic agriculture, agroecology, and other green topics, such as ecology, sustainability, environment, etc. The content is aggregated from a number of networked repositories, which is constantly increased, and it is appropriate for most types of education (school, university, vocational, and adult). For the creation of the metadata records, the Organic.Edunet metadata application profile has been used [9, 10], which is a version of the IEEE LOM standard modified in order to meet the requirements for the annotation of agricultural educational resources. For the classification of the resources, the Organic.Edunet ontology is being used and supports the semantic search function of the portal based on the semantic services module [11]. Despite the fact that access to the portal is free to all visitors, there are already more than 5,500 registered users, who are enabled to rate, review, and bookmark resources available through the portal, providing additional value to the existing metadata records. In the reference period for this study (1/1/2010–12/20/2012), the portal has received more than 156,000 visits and 478,000 page views from more than 128,000 unique visitors from 195 different countries.

Visits	Page views	Unique visitors	Number of countries	Browser languages
156,376	478,313	128,230	195	114

Table 1: Organic.Edunet portal statistics from the log files (Reference period: 1/1/2010–12/20/2012)

In order to meet the multilinguality requirements of its users, Organic.Edunet has worked toward the deployment of multilingual services and the aggregation of content in several languages. As regards the multilinguality aspects of the portal, its user interface has already been manually translated into seventeen languages by persons affiliated with the Organic.Edunet network. The actual content, as well as the corresponding metadata records, is available in several languages, mainly English but also Estonian, French, German, Greek, Hungarian, Norwegian, Romanian, Russian, Spanish, and Turkish. However, since Organic.Edunet is an active network, new collections are constantly added, providing in several cases content in additional languages or enriching the content in existing ones.

Portal language	Interface	Metadata application profile	Metadata	Available resources	Language related to countries studied?
English	√	√	10,883	8,522	Yes. South and West Africa, Asia
Greek	√	√	1,356	339	No
German	√	√	1,509	1,135	No
Spanish	√	√	599	99	Yes. Latin America
Estonian	√	√	1,241	311	No
French	√	√	174	3	Yes. North Africa
Turkish	√	√	-	1	No
Arabic	√	√	-	-	Yes. North Africa

Portal language	Interface	Metadata application profile	Metadata	Available resources	Language related to countries studied?
Hungarian	√	√	2,204	293	No
Bulgarian	√	√	-	-	No
Hindi	√	√	-	-	Yes. India
Norwegian	√	√	1,082	41	No
Romanian	√	√	1,249	272	No
Russian	√	√	1,012	27	No
Slovenian	√	√	2	1	No
Chinese	√	√	-	-	Yes. China
Italian	√	√	1	1	No

Table 2: Language and content statistics of the Organic.Edunet web portal, as of December 2012

The analysis of Table 2 shows that the portal is ready to support content in additional languages that are not currently represented in the pool of educational resources available through the portal, such as Arabic, Chinese, Hindi, and other languages, for which the translations of both the user interface as well as the metadata AP and the ontology have already been provided. However, due to the fact that the localization of both the user interface and the content is a time-consuming and costly process, the real needs of the users have to be carefully identified and studied before the localization takes place [12]. This work aims to identify the behaviour of users from specific developing countries, which may contribute to understanding the multilinguality needs of these users.

B. Use of the Portal in Developing Countries

According to the International Statistical Institute, "countries are divided into developed or developing according to their Gross National Income (GNI) per capita per year. Countries with a GNI of US$ 11,905 and less in 2010 are defined as developing" [13]. The United Nations Statistics Division provides an alternative, more abstract definition of developing countries:

> There is no established convention for the designation of "developed" and "developing" countries or areas in the United Nations system. In common practice, Japan in Asia, Canada and the United States in northern America, Australia and New Zealand in Oceania and Europe are considered 'developed' regions or areas. In international trade statistics, the Southern African Customs Union is also treated as developed region and Israel as a developed country; countries emerging from the former Yugoslavia are treated as developing countries; and countries of Eastern Europe and the former USSR countries in Europe are not included under either developed or developing regions [14].

Last but not least, the International Monetary Fund (IMF) provides a more financial-related aspect to the definition of developing countries: "The main criteria used by the World Economic Outlook (WEO) to classify the world into advanced economies and emerging market and developing economies are (1) per capita income level, (2) export diversification and (3) degree of integration into the global financial system" [15].

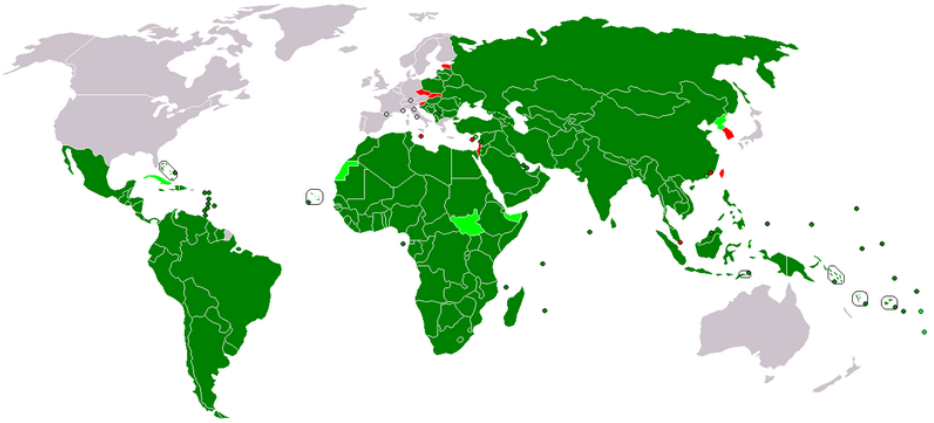

Figure 1: Map of developing countries within the scope of IMF (dark green) and outside the scope of IMF (light green) (Wikipedia: http://en.wikipedia.org/wiki/Developing_country)

Our approach was to use the list of developing countries, as provided by the International Statistical Institute [13], and focus on the countries from which a minimum number of 100 unique visits were recorded in the reference period. These countries were then organized according to their geographic region in order to allow easier management of the analytics. The results of this filtering can be seen in Table 3.

Region	Country	Visits	Page views	Unique visitors	Browser languages
Latin America and the Caribbean		4,091	7,952	3,678	
	Mexico	785	1,293	735	es
	Colombia	438	972	375	es
	Argentina	375	924	340	es
	Peru	319	550	282	es
	Chile	306	576	271	es
	Venezuela	164	271	155	es
	Ecuador	141	258	131	es
Northern Africa		1,774	2,949	1,664	
	Tunisia	1,281	1,908	1,208	fr, en, ar
	Egypt	272	504	251	en, ar
	Morocco	104	240	95	fr
Southern Africa		418	796	385	

	South Africa	361	680	336	en
Eastern Africa		1,042	2,117	923	
	Ethiopia	256	512	236	en
	Kenya	252	519	215	en
	Tanzania	227	448	209	en
Western Africa		499	871	469	
	Nigeria	329	522	307	en
	Ghana	107	196	102	en
Southern Asia		6,130	11,104	5,678	
	India	4,791	8,409	4,455	en
	Iran	542	1,306	472	en
	Pakistan	410	644	392	en
	Sri Lanka	150	282	135	en
	Bangladesh	109	217	101	en
Southeastern Asia		2,308	3,723	2,176	
	Philippines	776	1,160	738	en, fil
	Malaysia	453	700	426	en
	Indonesia	349	526	333	en, id
	Thailand	289	558	272	en
	Vietnam	210	401	193	en, vi
East Asia		1,567	2,817	1,400	
	China	648	1,143	573	zh-cn, en

Table 3: List of developing countries included in this study

The 17th Edition of the Ethnologue database (www.ethnologue.com), developed by Lewis et al. [16], provides information on languages spoken within each country and was used as a source of related information for the purposes of our analysis.

Latin America and the Caribbean: Latin America and the Caribbean provided a significant number of visits to the Organic.Edunet web portal. The countries from this region that provided 100 visits or more were Mexico, Colombia, Argentina, Peru, Chile, Venezuela, and Ecuador. The language spoken in these countries is Spanish, which is one of the languages supported by Organic.Edunet both at the content and

at the user-interface levels. According to the logs of the portal, the users coming from these countries used a Spanish version of their web browser.

Northern Africa: Tunisia, Egypt, and Morocco are the three countries from North Africa represented in this survey. The results showed that users from Tunisia, where the official language is Tunisian Arabic, mainly used web browsers in Arabic and French, along with the more generic English version. In Egypt, where the official language is Egyptian Arabic, the Arabic version was used. Last, despite the fact that the official language of Morocco is Arabic, it was noticed that the users mostly used the French version of the web browsers in order to visit the portal. This may be explained by the fact that the French language is also widely used in Morocco between scientists and researchers.

Southern Africa: Only South Africa provided more than 100 visits from the Southern Africa region. South Africa has eleven official languages: Afrikaans, English, Ndebele, Northern Sotho, Sotho, Swazi, Tswana, Tsonga, Venda, Xhosa, and Zulu; however, less than 2% of South Africans speak a first language other than an official one. In any case, the most widely used languages in South Africa are English and Afrikaans [17]. This explains the fact that the English user environment was used for accessing Organic.Edunet.

Eastern Africa: Ethiopia, Kenya, and Tanzania were the countries that provided more than 100 visits to the Organic.Edunet web portal. In all cases an English web browser was used for accessing the portal, probably due to the fact that there was no localized version of the browser in the languages spoken in these countries (Amharic and Swahili).

Western Africa: The Organic.Edunet web portal received a relatively high number of visits from Nigeria and Ghana. In Nigeria there are more than 500 languages spoken, while in Ghana there are almost 80. In both countries the official language is English, which might explain the fact that the visits were made with an English browser.

Southern Asia: As regards India, there are more than 100 languages spoken, while the official language is Hindi, and English is also widely spoken. In Iran the official language is Persian. In Pakistan the official languages are Urdu and English. Sri Lanka has two official languages, Sinhala and Tamil, while in Bangladesh the official language is Bengali. In all cases, the vast majority of the visits were made with an English browser.

Southeastern Asia: The official languages of Philippines are Filipino and English. Malay is the official language in Malaysia, and a related one, Indonesian, is the official language of Indonesia, among the almost 700 languages spoken. Thai and Viet are the official languages of Thailand and Vietnam, respectively. Only in the case of Philippines, Indonesia, and Vietnam were there visits in the official languages of these countries, while in the rest of the cases, the English version of the web browser was used by the visitors of the portal.

East Asia: China was identified as the only developing country from East Asia. The official language is Standard Chinese, and the visits received were made both in this language and in English.

III. METHODOLOGY

A. Research Problem and Questions

The aim of this work was to analyze the information related to the usage of the Organic.Edunet web portal by users from developing countries, in order to better understand the multilinguality use of the portal as well as possible issues that could be identified. Among the aspects studied are the language-related backgrounds of the users, the multilingual requirements of the users, and the way that the portal's multilinguality features affect the behavior of the users.

The reference period for this study is the one between the launch of the portal (1/1/2010) and near the end of the third year of the portal (12/20/2012), during which the status of the portal was established and a normal operation had been achieved for quite a long time. Therefore, the results of this analysis may be considered representative.

B. Variables Studied

The analysis described in this paper was based on previous work by Stoitsis et al. [2]. The different types of variables that needed to be studied in order to export the necessary information from the log files in this specific paper, along with the corresponding descriptions, may be found in Table 4.

Variable	Description
Visits (per day)	Visits per day for countries with full, partial (only interface), and no language support
Bounces (per day)	Bounces per day for countries with full, partial, and no language support
Page views (per day)	Page views per day for countries with full, partial, and no language support
Unique visitors (per day)	Unique visitors per day for countries with full, partial, and no language support
Average time on site (per visit)	Average time on site for users from countries with full, partial, and no language support
Average time on page (per visit)	Average time on page for users from countries with full, partial, and no language support
Page views (per visit)	Number of pages per visit from countries with full, partial, and no language support
Visits from search engines	Number of visits from search engines for each language
Search depth	Average number of pages that users view after using portal search function for each language
Total unique searches	Total number of unique searches using portal search function for each language
Number of key words used in portal search	Total number of key words used in portal search for each language

Table 4: A list of variables used in this study, as defined by Stoitsis et al. [2]

C. Methods and Tools

In order for us to be able to provide insight on the multilinguality-related aspects posed in the initial hypotheses of this work, we first needed to identify the corresponding variables that could provide the feedback required. In this direction, the following approaches were considered in each case, based on the previous work by Stoitsis et al. [2]:

1. **Language-related background of the users:** In order to obtain information related to the language-related background of the users, the following types of statistics were used:
   a) Descriptive statistics of the portal visits, such as total visits and page views
   b) Language and origin statistics: includes a number of indicators related to the location of the users, such as the country from which the visits were made, information about the language of the browser used for accessing the portal, and the official language of the country from where the user accessed the portal
   c) Combined statistics: refers to the combination of the two aforementioned types of statistics, which takes place when sufficient information is available
2. **Multilingual requirements of the users:** The following methods were used for accessing information about the multilingual requirements of the portal's users:
   a) Keywords used in generic search engines (e.g., Microsoft Bing, Google, and Yahoo!) for accessing the Organic.Edunet web portal

b) Keywords used in searches within the Organic.Edunet web portal using one of the available search functions

3. **Effect of the portal's multilinguality features on the behavior of the users:** This type of analysis was the most complex one, as it required correlation between different types of information. An example would be to correlate the number of visits from a specific country and the possibility that the official language spoken in this country is supported or not by the Organic.Edunet web portal.

The information required for the aforementioned types of analysis was mostly retrieved from the Google Analytics service (http://www.google.com/analytics), which has been set up for the portal since its launch, and therefore provides an overview of the corresponding indicators. In some cases, the use of Microsoft Excel was necessary in order to work on statistical formulas and correlations between different indicators.

IV. RESULTS

A. Language-Related Background of the Users

As regards the language-related background of the users, the available information both for the country of the user as well as for the language of the browser used for accessing the portal was used. The analysis of the results is presented per geographic region, as defined in the previous chapters. Some of the indicators for the reference period may be found in Table 5. For this reference period, the portal received visits from 195 countries all over the world, which resulted in a relatively high diversity of the indicators.

	Unique visitors/day	Bounces/day	Page views/day	Visits/day
World	118.29	85.10	441.25	144.26

Table 5: Basic indicators for the reference period

Table 6 provides an analysis of indicators related to portal-level statistics for the aforementioned regions in relation to the language of the browser used by the users when accessing the Organic.Edunet web portal.

Region	Browser language(s)	Unique visitors/day	Bounces/day	Page views/day	Visits/day
Latin America and the Caribbean	es	3.38	1.99	7.31	3.76
Northern Africa	en, fr, ar	1.53	1.27	2.71	1.63
Southern Africa	en	0.35	0.26	0.73	0.38
Eastern Africa	en	0.85	0.60	1.95	0.96
Western Africa	en	0.43	0.32	0.80	0.46
Southern Asia	en	5.22	3.90	10.21	5.63
Southeastern Asia	en, fil, id, vi	2.00	1.48	3.42	2.12
East Asia	cn	1.29	0.97	2.59	1.44

Table 6: Average values of portal-level statistics for the regions of the developing countries

As regards the language of the browser used for accessing the portal, an interesting point is that in several cases the vast majority of the users preferred the English version of their browsers instead of a localized version, if available. This was the case especially for African countries and might imply that these users were also able to make use of the English user interface of the portal as well as English educational resources available through the portal. This is also indicated by the lower indicator levels related to their activity in the portal.

The highest number of unique visitors per day that comes from the Southern Asia region is mainly influenced by the activity of the Indian users. In this region, the users used the English version of the browsers in order to visit the portal; therefore, we may also consider the possibility that they were not really affected by the fact that there was no content in Hindi, as they could probably use the available English resources. The level of the majority of the indicators that this specific region exhibited is more than double compared to the rest of the regions examined. In addition, increased values of the indicators compared to the rest of the regions analyzed in this work are provided by the Spanish-speaking Latin American countries. Despite the fact that the results may be affected by a variety of factors (such as the population of each country, the size of the community engaged in activities related to the context of Organic.Edunet, etc.), this may be partially explained by the fact that the portal supports Spanish both at the interface level and at the content/metadata level, facilitating the usage by Spanish-speaking users.

The same seems to apply for the visits and page views per day, as this metric appears to be higher in regions where the visits were made with browsers either in the local language or their English versions where English is widely spoken, such as Southern Asia. In all cases, the specific metrics seem to be lower in the African countries, where there was neither user interface nor content in the native languages.

B. Multilingual Requirements of the Users

After studying the linguistic backgrounds of the portal's users, we moved to the next level, which was mainly based on how users used queries, both localized and in English, and how often they used localized terms against the widely used English ones. We considered that the most appropriate way to study the multilingual requirements of the users would be to examine the search queries performed by the users, both for reaching the Organic.Edunet web portal as well as for retrieving results within the portal using one of the portal's search mechanisms. In both cases we focused on the languages identified in the previous paragraphs, based on the official languages spoken in the countries studied in this work and the languages of the browsers used by the users to access the portal.

As regards the queries in generic search engines that led to the Organic.Edunet web portal, a list of the 500 most frequently used terms out of the 44,485 in total was extracted from the logs of the portal and the language of each term was identified using the Google Translate API (https://developers.google.com/translate). The next step included filtering the languages so that only the ones that might have been used by users from the participating countries were kept; this left us with only 99 terms in English and Spanish, as the majority of the queries performed within the portal were in European languages that did not belong to developing countries (such as German, Greek, Estonian, Hungarian, Norwegian, Romanian, Russian, and Turkish). The metrics used in this analysis include the language of the terms, the number of visits made to the portal using these terms in generic search engines, the total number of terms identified in this language, and the average number of pages viewed after using a query with these terms.

Term language	Total number of visits	Total number of terms	Search depth
English	10,185	93	3.30
Spanish	498	6	1.37

Table 7: Information on the queries performed in generic web engines before reaching the Organic.Edunet web portal

Table 7 shows that the English terms used for accessing the portal lead to a higher number of pages viewed per visit compared to the Spanish ones. However, due to the small sample size of the latter and the fact that English was used by users from developed countries, no safe outcomes could be extracted.

The most widely used queries performed by users, which eventually led them to the Organic.Edunet web portal, are presented in Table 8.

Query term	Term language	Unique queries
organic edunet	English	2243
organic.edunet	English	1386
edunet	English	1126
pet bottle compactor	English	461
budapest t2	English	431
organic edunet portal	English	380
man and natural environment	English	247
camelina sativa	Latin	214
recursos secundarios	Spanish	205
agbios	English	200
how a corn plant develops	English	149
www.organicedunet.eu	English	113
organicedunet	Spanish	103
portal organic edunet	Spanish	95
educational scenario	English	92
cicer arietinum	Latin	82
natural and manmade environment	English	82

Table 8: The most popular queries of users for reaching the Organic.Edunet web portal

Table 8 shows that the English terms are dominating the list of most popular queries made for accessing the portal. Apart from the few Spanish ones, there were also queries made using the scientific names of specific plants in Latin.

As regards the queries used within the portal by the users for retrieving related resources, we followed a similar approach in order to record the corresponding indicators, filtering the first 500 terms out of a list of total 5,123 terms identified in the logs. This resulted in 108 terms used in languages that might have been used by users from the participating countries. The metrics used in this analysis include the language of the terms, the number of queries made using these terms within the Organic.Edunet web portal, the total number of terms identified in this language, and the average number of pages viewed after using a query with these terms. The results can be found in Table 9.

Term language	Unique queries	Total number of terms	Search depth
English	2,547	88	2.86
Spanish	426	19	6.21
French	9	1	3.67

Table 9: Information on the queries performed in the Organic.Edunet Web portal

As expected, the queries with English terms were the most popular, but at the same time they led to fewer pages viewed per visit. On the other hand, queries in Spanish and French were smaller in quantity but kept the users more engaged in the portal. In addition, it is once more obvious that no queries in Arabic, Chinese, or Hindi were identified in the 500 most popular ones, a fact which might be explained by the lack of content in these languages in the Organic.Edunet web portal.

The most popular queries within the Organic.Edunet web portal in the three languages identified are presented in Table 10. We will see that the terms used in this case are more specific to the context of Organic.Edunet and that the English versions dominate the list.

Query term	Term language	Unique queries
School	English	369
Scenario description	English	265
School garden	English	212
Handbook scenario implementation university level	English	192
Handbook scenario implementation school level	English	179
Climate change	English	122
Organic garden (EA)	English	82
Tomate	Spanish	72
Organic	English	67
Patata	Spanish	64
Tomato	English	51
Comparing organic farming (BG/BRG Schwechat)	English	43
Control de plagas	Spanish	42
Leche	Spanish	41
Introduction to OA (USAMVB)	English	39
Lechuga	Spanish	39
Control plagas	Spanish	37

Table 10: The most popular queries in the Organic.Edunet web portal

The fact that among the widely used English terms there are also six terms in Spanish indicates the high activity of the Spanish-speaking users within the portal. We may only assume that this would be the case in other instances as well, such as the French-speaking users, if there were French content available within the portal.

C. Effect of the Portal's Multilinguality Features on the Behavior of the Users

In this section, the analysis of the behavior of the users within the portal is described, based on the multilinguality features and aspects of the portal. In order to facilitate the analysis and be able to export meaningful results, we decided to limit the analysis of the indicators used to only two cases:

a) Countries with full language support, including countries that use languages in which both the user interface of the portal as well as the metadata and/or resources are available

b) Countries with only interface support, including countries that use languages in which the user interface of the portal is available but there are no resources or metadata records

A case study for the first category is the group of Latin American countries, while for the second one it is regions like Northern Africa and Eastern Asia.

In order for us to identify the impact that the availability of multilingual services has on the use of the portal by users belonging to one of the two aforementioned categories, we used two types of indicators:

a) **Portal level:** The portal-level indicators include unique visitors per day, bounces per day, page views per day, and visits per day.

b) **User level:** The user-level indicators include average time on site, average time on page, and pages per visit.

The combination of the outcomes from these two kinds of indicators was also used by Stoitsis et al. [2] and has provided a set of interesting results. Based on this approach, we also investigated the relation between the activity of the users in the portal and the use of the multilinguality services.

Table 11 provides portal-level statistics for the countries with language support. Countries that provided visits from visitors with browsers in English are included in this category. Wide variations can be identified between the different regions. However, we could assume that the higher number of visitors per day and page views per day in Latin American countries is due to the full support of their native language. On the other hand, despite the fact that in Southern Asia the vast majority of the visits took place with an English browser, English is not the official language of these countries, but still the number of both visits and visitors per day was higher than in the rest of the regions. The same might apply in the case of the African countries, where visits to the portal were made using an English browser as well; however, English is not the official language of these countries; therefore, no material in the native languages of the users could be retrieved.

Country/Territory	Unique visitors/day	Bounces/day	Page views/day	Visits/day
Latin America and the Caribbean	3.38	1.99	7.31	3.76
Southern Africa	0.35	0.26	0.73	0.38
Eastern Africa	0.85	0.60	1.95	0.96
Western Africa	0.43	0.32	0.80	0.46
Southern Asia	5.22	3.90	10.21	5.63
Southeastern Asia	2.00	1.48	3.42	2.12
Average	**2.03**	**1.43**	**4.07**	**2.22**

Table 11: Average values of portal-level statistics for the countries with language support

Table 12 provides the corresponding information for Northern Africa and Eastern Asia, whose official languages are Arabic and Chinese, respectively, and which are only supported by Organic.Edunet at the user-interface level. The lack of content in Arabic and Chinese might be one of the reasons that led to lower activity in the portal compared to the previously described regions.

Country/Territory	Unique visitors/day	Bounces/day	Page views/day	Visits/day
Northern Africa	1.53	1.27	2.71	1.63
Eastern Asia	0.53	0.40	1.05	0.60
Average	**1.03**	**0.84**	**1.88**	**1.12**

Table 12: Average values of portal-level statistics for the countries with interface support

In all cases the indicators exhibit significantly lower levels compared to the previous set of regions, and in most cases these levels are almost half of those of the previous set of regions. This outcome could also be considered as a factor explaining the lower levels of participation of users from these regions, as they would prefer to retrieve content in their own language.

Moving to the user-level statistics for the countries with language support, Table 13 shows that users from Eastern and Western Africa tend to spend more time in the portal and visit a rather higher number of pages per visit despite the fact that there is no content in their languages. It should be noted that in both cases the English versions of the web browsers were used for accessing the portal, which might imply that these users were also capable of using the English content available through the portal.

Country/Territory	Average time on site (sec)	Average time on page (sec)	Pages per visit
Latin America and the Caribbean	77	91	1.83
Southern Africa	97	107	1.90
Eastern Africa	117	113	2.03
Western Africa	110	148	1.75
Southern Asia	73	90	1.81
Southeastern Asia	66	107	1.61
Average	**90**	**109**	**1.82**

Table 13: Average values of user-level statistics for the countries with language support

Table 14 provides the corresponding information for the regions in which the official languages are Arabic and Chinese, respectively, languages in which there are no metadata or resources in the portal. Despite that fact, there are no significant differences between these regions and the ones belonging to the category of those whose languages are represented with metadata in the Organic.Edunet web portal. Nevertheless, it should be noted that the difference in the recorded values of the indicators studied for the two regions of this category are significant.

Country/Territory	Average time on site (sec)	Average time on page (sec)	Pages per visit
North Africa	47	71	1.66
Eastern Asia	118	146	1.76
Average	**83**	**109**	**1.71**

Table 14: Average values of user-level statistics for the countries with interface support

V. DISCUSSION

The objective of this study was to identify the multilinguality aspects of the Organic.Edunet web portal that could potentially influence the behavior of its users, based on the analysis of the log files of the portal. The language-related background of the users provided additional information on this aspect and allowed for more objective results. In the vast majority of the cases, the English version of the web browser was used for accessing the portal. This fact raised barriers to identifying the behavior of the users who only used their native language, because those users would be prohibited from accessing and using a web portal that was not translated into their language and that did not provide educational material, or at least metadata, in the their language. In the case of tutors, researchers, and other educated users of the portal, it is a common practice to use the English version of the portal rather than the one in the native language of the user, in order to ensure that the actual information is used and to avoid any omissions and/or errors that may occur in the translation of the information.

The results of our analysis showed that English was the dominant language used by the users, both for accessing the portal by searching in generic web search engines and for searching content within the portal. This was expected due to the universal acceptance of the English language. This massive use of English terms did not allow less frequently used terms in other languages to be studied and discussed in the context of this work. In addition, our research also showed that in several cases, users used the English version of their web browsers for accessing the portal instead of a localized version in their native language, which also explains why the English language was widely used in our study. The identification of the queries performed both outside and inside the portal also revealed that the majority of the terms were in several European languages that were not studied in the context of this work. However, this rendered the identification of potential queries made in one of the languages studied, such as Arabic, Hindi, and Chinese, a difficult task, as these queries were probably at the lower pages of the search results. In the cases mentioned in this study, it was interesting for one to see that the queries that were

performed in languages other than English (namely Spanish and French) led to more page views per visit, which is also an indicator that users get more engaged in the portal when they are able to identify content in their native language.

Portal-level statistics indicated that the activity of the users in the regions where countries had language support was significantly higher—more than double the numbers in almost all other cases. This could be an indicator connecting the multilingual services of the portal with the native language of the users. However, the user-level statistics did not provide a similarly clear difference between the two types of regions, as the differences noticed could not be considered as significant. In cases such as India, Arabic-speaking countries, and China, the possibility of increased activity in the portal if there was localized material in the corresponding languages could be considered based on the outcomes of this study but could not be confirmed due to lack of required details in each case.

VI. CONCLUSIONS

The work presented in this paper describes a part of the multilinguality aspects of a learning portal devoted to organic agriculture and agroecology and tries to interpret the behavior of its users from the view of the use of these aspects. It has already been discussed that the availability of non-English documents on the web is rapidly growing, and users need to find a way to access them using an interface localized in their native language [18]. We tried to organize the aspects in a way that would make sense both for the optimal identification of the context of the study as well as for the explanation of the results. However, we should note that this work was challenging due to the wealth of information that needed to be addressed. For this reason, and because the information related to each country individually was difficult to allow its management at a country level in a single publication, it was decided that the corresponding regions should be used instead, facilitating the grouping of related information. However, we acknowledge that this could potentially lead to loss of details in the way that the results are presented.

As a next step, we aim to focus on specific regions of interest, based on the outcomes of this study. Some of the regions that could be more thoroughly examined could be Northern and Eastern Asia, Northern Africa, and Latin America. These regions include countries with widely spoken languages, such as Arabic, Chinese, and Hindi, and therefore exhibit a high interest for a multilingual portal. Future studies could focus on the countries included in the aforementioned regions and identify differences that may exist in the behavior of users originating from these countries. The outcomes of these future studies could be used for identifying the cases where the effort and cost of providing a localized version of the portal would be justified by the increased use.

There are publications that confirm that there is a globally increased demand for education, while at the same time there is a growing gap between the demand and the availability of educational sources [19]. This demand is not constant but varies between different countries and types of education [20]. This gap may be also due to multilinguality issues identified in the learning portals, an issue that has to be investigated more thoroughly in upcoming related studies.

VII. ACKNOWLEDGEMENTS

The work presented in this paper has been funded with support by the European Commission and, more specifically, the project CIP-ICT-PSP-270999 “Organic.Lingua: Demonstrating the Potential of a Multilingual Web Portal for Sustainable Agricultural & Environmental Education” of the ICT Policy Support Programme (ICT PSP).

VIII. ABOUT THE AUTHORS

Vassilis Protonotarios holds a diploma in Agricultural Sciences from the Agricultural University of Athens (2001) and an MSc (2003) and PhD (2008) in Agricultural Biotechnology from the same university. He has professional experience in organic agriculture as an ex-officer of the Greek Ministry of Rural Development and Food, Directorate of Organic Agriculture. He has been working with educational

metadata, digital collections/repositories, and learning portals since 2009. He is a scientific associate of Agro-Know Technologies, the Greek Research and Technology Network, and recently the University of Alcala (Spain).

Giannis Stoitsis received a Diploma of Electrical and Computer Engineering from the Aristotle University of Thessaloniki in 2002, and MSc and PhD degrees in Biomedical Engineering from the University of Patras in 2004 and 2007, respectively. In the context of this study, Giannis took a place as a member of (and was funded by) the Information Engineering Unit of the University of Alcala. He has been previously affiliated with the Biomedical Simulations and Imaging (BIOSIM) Laboratory of the National Technical University of Athens (NTUA), Greece. His research interests include data repositories, Semantic Web technologies, and biomedical image processing.

Kostas Kastrantas holds a diploma in Electrical and Computer Engineering from the Polytechnic School of Aristotle University of Thessaloniki, Greece (2002). He also holds an MSc in eLearning from the University of Piraeus, Greece (2007). He has strong experience in developing web-based applications, and his main interests focus on online ontology-based and metadata applications. He has a long working record in research and academic settings with teams specializing in learning technologies.

Salvador Sanchez-Alonso holds a PhD in Computer Science (2005) and a degree in Information and Library Science from the University of Alcala (2010). He is an expert in learning object metadata and the application of information and semantic technologies to education. Since 2005, he has been part of the Computer Science Dept. of UAH and a member of the Information Engineering Research Unit. He is involved in the coordination of the CIP PSP Organic.Lingua and LdV Organic.Mednet projects.

IX. REFERENCES

1. **Siemens, G**. What Are Learning Analytics? *Elearnspace*, August 25, 2010. http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/
2. **Stoitsis, G., Manouselis, N., and Sanchez-Alonso, S.** Data Set Requirements for Multilingual Learning Analytics. *International Journal of Technology Enhanced Learning (IJTEL)* 4(1/2): 47–66 (2012). ISSN: 1753-5255.
3. **Greller, W. and Drachsler, H.** Translating Learning Into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society* 15(3): 42–57. http://ifets.info/issues.php?id=56
4. **Siemens, G. and Baker, R.** Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In LAK '12 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM: New York, NY, 252–254, 2012. ISBN: 978-1-4503-1111-3
5. **Verbert, K., Manouselis, N., Drachsler, H., and Duval, E.** Dataset-Driven Research to Support Learning and Knowledge Analytics. *Journal of Educational Technology & Society* 15(3): 133–148 (2012). http://ifets.info/issues.php?id=56
6. **Ferguson, R. and Buckingham Shum, S.** Learning Analytics to Identify Exploratory Dialogue Within Synchronous Text Chat. In *LAK 2011: Proceedings of the 1st International Conference on Learning Analytics and Knowledge 1st International Conference on Learning Analytics & Knowledge*, ACM: New York, NY, 99–103, 2011. ISBN: 978-1-4503-0944-8.
7. **Sampson, D. and Manouselis, N.** A Flexible Evaluation Framework for Web Portals Based on Multi-Criteria Analysis. In A. Tatnall, ed. *Web Portals: The New Gateways to Internet Information and Services,* 185–211. Idea Group Inc.: Hershey, PA, 185–211, 2005.
8. **Palavitsinis, N., Protonotarios, V., and Manouselis, N.** Applying Analytics for a Learning Portal: The Organic.Edunet Case Study. In *LAK 2011: Proceedings of the 1st International Conference on Learning Analytics and Knowledge 1st International Conference on Learning Analytics & Knowledge*, ACM: New York, NY, 99–103, 2011. ISBN: 978-1-4503-0944-8

9. **Palavitsinis, N., Ebner, H., Sanchez-Alonso, S., and Manouselis, N.** Using e-Learning Technologies and Standards to Make Educational Content Available: The Organic.Edunet Approach. *Journal of Information Technology in Agriculture* 4(1) (2011).
10. **Palavitsinis, N., Manouselis, N., and Sanchez-Alonso, S.** Evaluation of a Metadata Application Profile for Learning Resources on Organic Agriculture. *Metadata and Semantic Research - Communications in Computer and Information Science* 46: 270–281 (2009). DOI: 10.1007/978-3-642-04590-5_26
11. **Manouselis, N., Kastrantas, K., Sanchez-Alonso, S., Cáceres, J., Ebner, H., and Palmer, M.** Architecture of the Organic.edunet Web Portal. *International Journal of Web Portals (IJWP)* 1(1): 71–91 (2009). http://kmr.nada.kth.se/papers/Misc/IJWP_OrganicEdunet_final.pdf
12. **Gazzola, M. and Grin, F.** Assessing Efficiency and Fairness in Multilingual Communication: Towards a General Analytical Framework. *AILA Review*, 20: 87-105 (2007). ISSN: 1570-5595
13. **International Statistical Institute.** Developing Countries (2012). http://www.isi-web.org/component/content/article/5-root/root/577-developing2012
14. **United Nations Statistics Division.** Composition of Macro Geographical (Continental) Regions, Geographical Sub-Regions, and Selected Economic and Other Groupings, as revised on 10/17/2008. Retrieved 01/15/2013 from http://unstats.un.org/unsd/methods/m49/m49regin.htm
15. **International Monetary Fund.** World Economic Outlook (WEO) – Frequently Asked Questions (FAQ) (2013). http://www.imf.org/external/pubs/ft/weo/faq.htm#q4b
16. **Lewis, P., Simons, G., and Fennig, C., eds.** *Ethnologue: Languages of the World, Seventeenth Edition.* SIL International: Dallas, Texas, 2013. Online version: http://www.ethnologue.com.
17. **Pretoria Statistics South Africa.** *Census 2011: Census in Brief.* 23–25, 2012. ISBN: 978-0-6214-1388-5
18. **Zhou Y., Qin, J., Chen H., and Nunamaker, J.F.** Multilingual Web Retrieval: An Experiment on a Multilingual Business Intelligence Portal. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05)*, January 3–6, 2005.
19. **Geith, C., Vignare, K., Bourquin, L., and Thiagarajan, D.** Designing Corporate Training in Developing Economies Using Open Educational Resources. *Journal of Asynchronous Learning Networks* 14(3): 3–12 (2010).
20. **Geith, C. and Vignare, K.** Access to Education With Online Learning and Open Educational Resources: Can They Close the Gap? *Journal of Asynchronous Learning Networks* 12(1): 105–126 (February 2008).
21. **Geith, C., Vignare, K., Bourquin, L., and Thiagarajan, D.** Designing Corporate Training in Developing Economies Using Open Educational Resources. *Journal of Asynchronous Learning Networks* 14(3): 3–12 (2010).