Multilingual Knowledge Management

Daniel E. O'Leary

3660 Trousdale Parkway, University of Southern California, Los Angeles, CA 90089-0441
oleary@usc.edu

Words = ~ 7200

Abstract. Although there has been substantial research in knowledge management, there has been limited work in the area of multilingual knowledge management. The purpose of this paper is to review and summarize some of the existing and supporting literature surrounding the emerging field of multilingual knowledge management. It does that by reviewing recent applications from multiple fields and the presentation of multilingual information. The paper uses a theory about knowledge management and also examines supporting literature in translation, collaboration, ontologies and search.

Keywords: multilingual knowledge management; multilingual collaboration; multilingual content presentation; multilingual theory; multilingual ontologies and vocabularies.

1 Introduction

All types of organizations are affected by a multilingual requirement, particularly in digital environments. One of the key emerging issues associated with multinational companies and e-government is multilingual knowledge management. As those organizations face the need to provide digital knowledge resources, they also face demands for presenting multilingual digital resources as broader bases of Internet users look for digital "e-solutions." Accordingly, organizations need to determine to what extent they will provide multilingual or single language knowledge resources and address issues such how to present multilingual knowledge resources.

In the early days of the Internet, virtually all of the search engines and content was English (e.g., Peters and Sheriden 2000). However, in many settings that has now changed as users from all over the world, using many languages, are using the Internet. It would be too costly for firms to ignore large populations of different language speaking customers and vendors. Similarly, governments must address citizens and interested parties with different cultural and language backgrounds. As a result, global companies and governments at all levels, that must provide solutions to a wider range of users, are focused on providing multilingual capabilities and supporting multilingual knowledge management.

1.1 Purpose of This Paper

Although the trend toward supporting multilingual corporate and government requirements is undeniable, unfortunately, there is only limited literature providing an analysis of multilingual knowledge management system capabilities and applications. Most of the previous work on knowledge management has ignored multilingual issues. As a result, the purpose of this paper is to review what is emerging as a literature of multilingual knowledge management. Although to-date there has been limited research of "multilingual knowledge management," per se, we also will

examine some of the foundations and applications that are coalescing into an emerging field. More mature areas such as machine translation, are worthy of studies of their own and thus are out of the scope of this paper.

Even the notion of what is "multilingual," appears to have multiple interpretations. In the literature, multilingual can refer to anything that involves a "multilingual user interface" to "multilingual content" to both. Issues such as "how to do multilingual search," or "what is an ontology in a multilingual environment" need to be assessed and require further research. As a result, this paper attempts to structure multilingual knowledge management.

1.2 Plan of This Paper

This paper proceeds in the following manner. Section 1 has provided an introduction and motivation for the paper. Section 2 briefly reviews knowledge management, what it means for a system to be multilingual and some of the costs and benefits of multilingual knowledge management. Section 3 summarizes some of the research and applications of multilingual knowledge management. Section 4 discusses issues in the presentation of multilingual resources. Section 5 summarizes some of the limitations of machine translation. Section 6 analyzes some recent research from multilingual collaboration. Section 7 investigates categorization and definition of knowledge using multilingual ontologies and vocabularies. Section 8 analyzes some issues in multilingual search. Section 9 reviews some of the emerging research issues, while section 10 summarizes the paper.

2 Knowledge Management and Multilingual Costs and Benefits

This section briefly reviews some basic notions of what it means to be multilingual, knowledge management, and what are some of the costs and benefits of multilingual systems.

2.1 Multilingual

What it means to be "multilingual" appears to occur along a spectrum. As noted by Rozic-Hristovski et al. (2002), at one extreme, multilingual means being able to select a web portal interface language. From that perspective, multilingual is almost reduced to a presentation issue. At the other end of the spectrum, not only the interface but also the resources are available in multiple languages and the links to those resources are multilingual. For example, for Peters and Sheridan (2000, p.52) multilingual refers to "…accessing, querying and retrieving information from collections in any language …." In this latter case, multilingual generally refers to content and information about the content, and connecting the user with specific aspects of the content.

As part of the user interface, multilingual also can refer to the language used to do general communication with the user of a knowledge management system, as part of multilingual presentation. As an example, a "customer survey" was made available in multiple languages (figure 1) for a United Nations agency FAO (Food and Agricultural Organization).

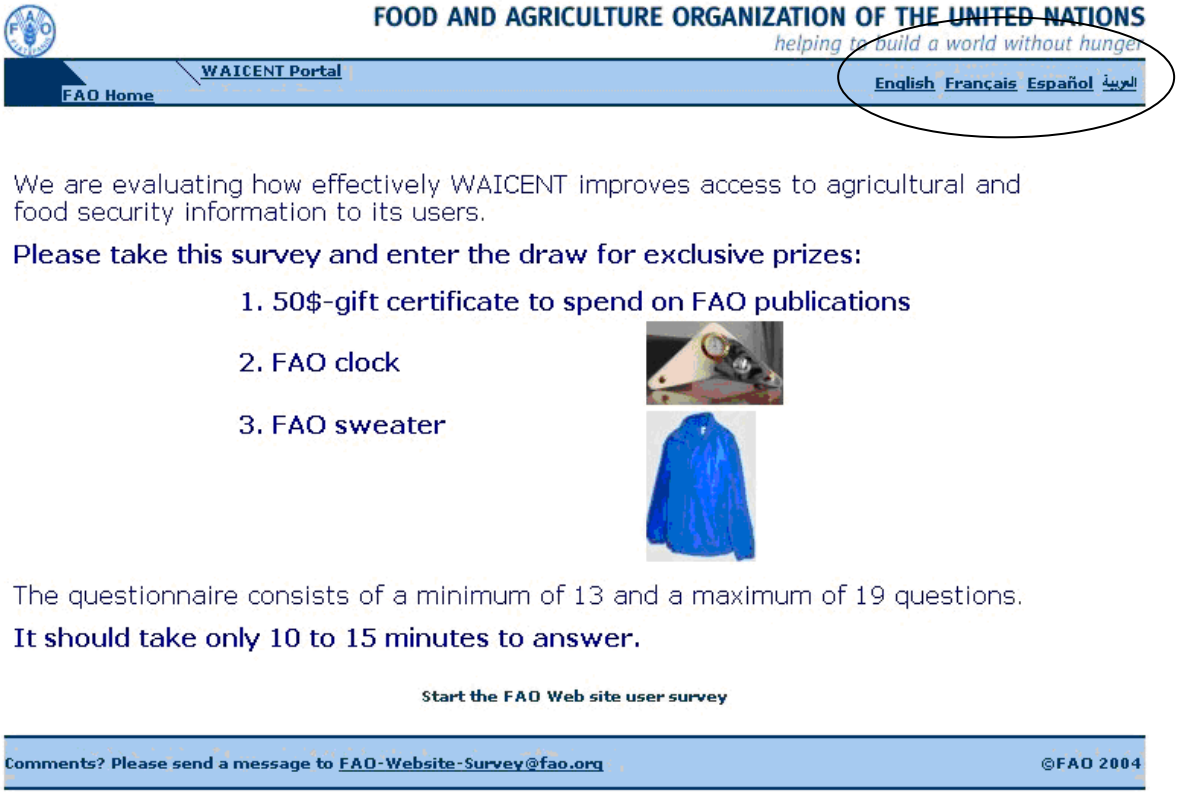

Figure 1
Multilingual Customer Survey

2.2 Knowledge Management

One approach to facilitating categorization of knowledge management capabilities was presented in O'Leary (1998a) and extended by O'Leary (2007a). Knowledge management systems have three primary capabilities. While providing appropriate knowledge "content," a knowledge management system may need to "convert" content to other languages and "connect" users to other users and knowledge resources (see figure 2).

"Content" includes a broad range of resources, such as knowledge about how to solve particular problems or information about particular products or other general information capabilities. Other critical content can include ontologies used to structure and search knowledge, to facilitate communication and to connect knowledge. As a result, along a dimension of (*multilingual) content*, organizations can have the capability to provide single language knowledge resources ("the official language is English") or they can provide content in multiple languages. In between those two extremes, firms can build multilingual presentation interfaces to single language or multilingual content.

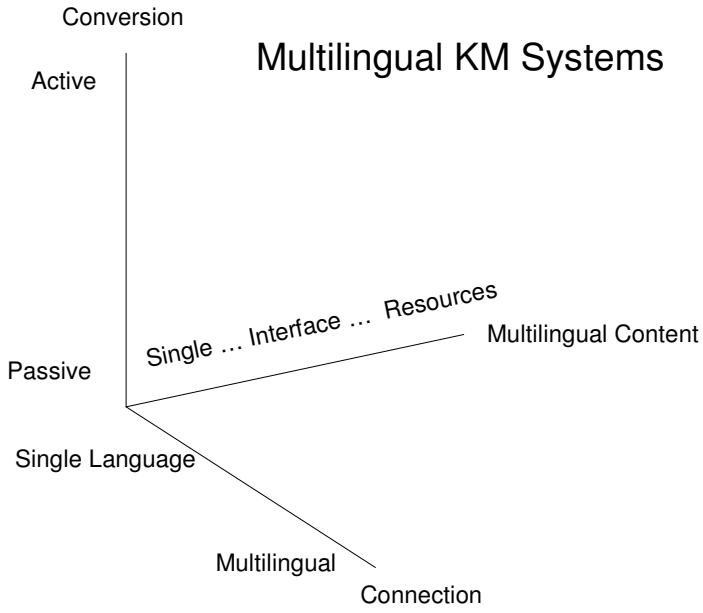

Figure 2
Categorizing Multilingual KM Systems

Individual knowledge needs to be converted to group available knowledge, and data and text need to be converted to usable knowledge, not to be lost in the piles and piles of data and text that are available. In the case of multilingual systems, organizations can be located along the spectrum of the *conversion* of knowledge resources to multilingual knowledge resources. Organizations can be completely passive or actively "convert" knowledge resources to multiple languages. In a multilingual system, knowledge resources can be converted from one language to others, e.g., using collaboration and translation devises.

Further, individuals need to be "connected" to knowledge resources and other people. Knowledge resources need to be searchable and links between appropriate knowledge islands need to be established. Further, connecting knowledge to other knowledge also must consider language, since in general; knowledge in different languages cannot be consumed by all users. *Connections* between people and other people, people and resources or resources and other resources, for example search or established links, or people and other people, can be based on a single language or use multiple languages. However, ultimately, search needs to provide useful connections in the language(s) appropriate for the user.

As an example, a firm that maintains a web presence in English and Spanish, by providing the same knowledge resources in both languages, and does that by actively generating the multilingual content themselves will be at the end of all three spectrums.

Historically, researchers (O'Leary 1998a) have not considered the multilingual aspects of those capabilities. As a result, from a *content*, *connection* and *conversion* perspective, some characterizations of knowledge management have been historically underspecified by not considering multilingual implications.

In a contemporary multilingual knowledge management system, content, conversion, and connection typically are accomplished using presentation, translation, collaboration, categorization/definition and search, among other activities. These activities and capabilities are summarized in figure 3, along with their most frequent interactions.

	Content	Converting	Connecting
Presentation			X
Translation		X	X
Collaboration	X	X	X
Categorization /Definition	X		X
Search			X

Figure 3
Knowledge Management Capabilities and Functions

Knowledge management content must be *presented* in a format accessible enough so that users are aware of, and can access those multilingual capabilities. For example, in a multilingual environment, users need to be able to readily change presentation language for the content from one language to another based on their native language. In this way, the user is "connected" to the "content" through the presentation. Knowledge management systems may convert content in a number of different ways. However, with multilingual systems, *translation* provides one of the key knowledge sources, using content in one language to convert to content in another language.

Collaboration can be a source of content in any knowledge management system. Further, collaboration can help convert knowledge resources from one language to another. Finally, collaboration facilitates connecting, for example, one person to another. Categorization/Definition

is most likely to be accomplished in contemporary knowledge management systems using ontologies. Ontologies have emerged as important tool in multilingual knowledge management systems providing structure and definition to knowledge content that also can be used for search. However, it also is likely that those ontologies need to be connected, so that changes in one language ontology, result in changes in other language versions, and so the ontology changes seamlessly as the user moves from one language to another. Search provides a major function of connecting users to information, and capturing and indexing related information.

2.3 Costs and Benefits of Multilingual Systems

Multilingual capabilities are not cost free, and there are a number of potential costs and benefits of multilingual systems. First, there are maintenance costs associated with generating all knowledge resources in multiple languages. Not only are there the costs of normally generated knowledge assets in a multilingual environment, but knowledge resources must be translated to meet the needs of other users. Second, if the languages are not translated correctly or completely there can be even larger costs. For example, imagine if laws were incorrectly translated, and that inconsistent translation caused people to act in a particular way, while relying on the inconsistent translation. Third, with a multilingual system, generally, translating from one language to another takes time. If there is emergency information, many constituencies can suffer if critical information is not posted until it is available in multiple languages. Consider for example, information about a potential bird flu pandemic. If information was held up while it was translated into another language, such an action could result in substantial human destruction. Thus, time wasted by not publishing all ready translated information can be high. Fourth, however, if knowledge resources are translated into one language but not another, then users may perceive a bias for those constituents of the first language over those of the second. In the course of politics or with consumer groups that alienation could be quite costly. Accordingly, when to make resources available in a multilingual environment is not clear. Fifth, costs of presenting knowledge resources in one language may be less costly than other languages because of the availability and quality of translation capabilities. As a result, multilingual capabilities are subject to resource constraints and considerations. Sixth, movement to multiple languages increases the complexity of the knowledge management system. If a single language is added and that language is keep completely separate, there will be twice the resources, etc. But now imagine the interaction between each of the knowledge resources. The number of potential links between knowledge resources can explode, increasing complexity.

However, multilingual systems potentially have a number of benefits. First, by putting information into a single language, the number of users of the web material are immediately limited, no matter what the language is. Thus having resources available in multiple languages provides access to a greater base of users for important issues (e.g., figure 4). Second, "transparency" can be increased by providing resources in more than one language. Providing materials in other languages opens up the web pages to many other potential users. Third, along with transparency, multilingual capabilities potentially generate a greater trust of the organization. Rather than hiding behind any one language a multilingual appearance provides greater access to knowledge about the organization. Finally, multilingual capabilities likely show a user centric view that attempts to provide the appropriate information to the public. When a user sees an organization that provides access and content in multiple languages, it provides a view that the organization is "concerned" about the user.

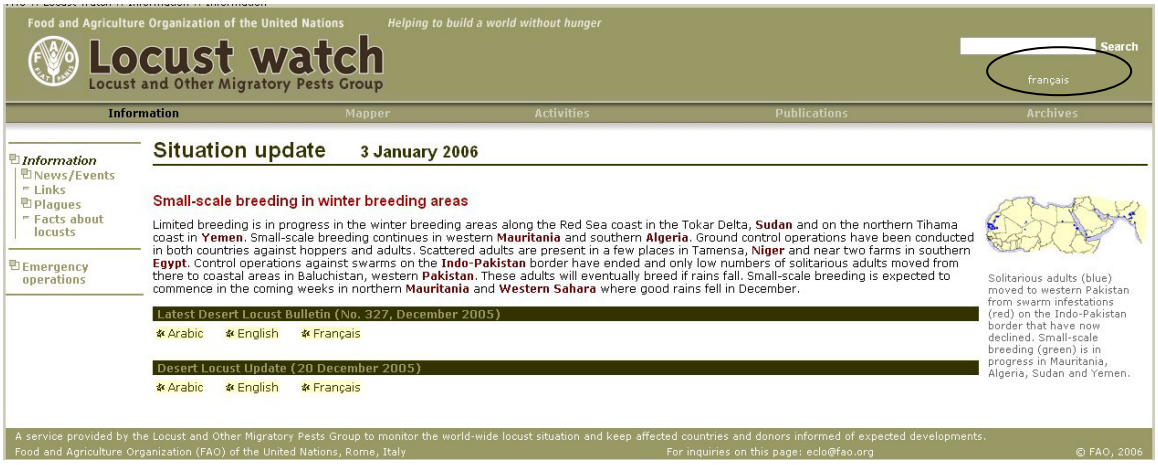

Figure 4
Multilingual Situation Update

3 Multilingual Knowledge Management Applications

In many ways, practice and the development of applications are leading theory in terms of addressing multilingual knowledge management. Organizations need to address multilingual issues now and cannot wait for theory to be developed. In particular, multilingual knowledge management is receiving attention in practice in a number of settings, including multinational firms, e-business, e-government, libraries, medicine and other international organizations.

3.1 Knowledge Management at a Multinational Firm

In a multinational firm, it probably is inevitable that there will be a demand for knowledge resources in multiple languages. One case study (O'Leary 2007b, p. 1142) of the large professional services firm KPMG found the firm concerned that different cultures, business cultures, and different languages stood in the way of the firm being a global firm. As a result, that firm ultimately kept "corporate" knowledge resources in a single language (English). However, that same firm apparently also allowed local offices to put additional servers onto the knowledge management network, and information on those servers could be placed in the originating and native languages.

Enterprises can take the position that they want a single voice / language between their employees. However, they are still likely to want to make resources, such as proposals to customers, etc. available in the native languages of their customers. Resources need to be available in those native languages so that those resources can be messaged by workers for their customers that use those native languages. As a result, ultimately, multiple languages need to be accommodated, even in a firm where there supposedly is a single language.

3.2 E-Business Knowledge Resources

Although individual firms may be able to declare use of a single language internally, their external face generally would need to account for each language used by major groups of customers. Another study (O'Leary 2007c) investigated the web presence of twenty-five of the largest firms in the world, the so-called "Fortune 25." That research found that roughly one-half of the firms had multilingual external web presences, although only one of the twenty-five firms had more than two languages. O'Leary found that the existence of a multilingual presence apparently was generated by a number of factors. First, the dominant language was English, with only one of the Fortune 25 not providing resources in that language. Second, there was a "head quarters" effect. If there was a second language, it was likely to be the language of the country in which the head quarters of the firm was located. Third, there appeared to be an "interested parties" effect. For example, recently there has been substantial legislation in the United States aimed at regulation of enterprises. As a result, some firms disclose registration information, but only in English in order to meet the needs of the government. Fourth, there appears to be an "industry" effect. Firms in the same industry tended to disclose information in the same language as the others in the same industry. Fifth, only one organization wrapped language with culture, with the label "Hispanic." As a result, multilingual continues to refer primarily to language, and not the cultures that they typically bring with them.

3.3 E-Government Knowledge Resources

E-government is the provision of knowledge resources and the capability to perform governmental processes on-line, typically in a web-based digital environment. There are a number of reasons why those e-government knowledge resources also need to be multilingual. For example, a country may have multiple official languages, e.g., in Belgium, French, Dutch and German are official languages. Further, a government may be part of a larger community. For example, countries in the European Union also have their own governments. As part of that larger community, countries may provide information in alternative languages beyond their official languages to facilitate "transparency."

What knowledge resources are provided by the governments in multilingual environments varies by context and government. For example, O'Leary (2007) studied the multilingual disclosures of the United States government. In addition to a "Spanish" option offered on many of its web pages, information about a broad range of activities were found to be disclosed in multiple languages, as summarized in table 1. "Family, Health and Safety" and "Visitors to the US" were the most frequently provided information. As a result, it appears as though there is a "Category of Disclosure" effect with significantly more disclosures in different categories. In addition, there was a "country effect," with differential multilingual resources provided in Korean, Russian and French.

	Busi-nesses	Civil Rights /Laws	Employ-ment	Family, Health Safety	Money/ Benefits	News	Visitors to US	Total
Arabic	1	1	1	1			1	5
Armenian					1	1		2
Cambod'n		1	1	1				3
Cantonese							1	1
Chinese	1	1	1	1	1		1	6
Dutch							1	1
Farsi			1		1			2
French	1	1	1	1	1	1	1	7
German				1			1	2
Greek					1	1		2
Haitian		1	1	1	1	1		5
Hebrew				1			1	2
Hindi			1	1		1		3
Hmong			1	1			1	3
Italian			1	1	1			3
Japanese	1			1			1	3
Korean	1	1	1	1	1	1	1	8
Laotian	1			1			1	3
Polish			1	1	1		1	4
Portug'se			1	1	1	1	1	6
Punjabi							1	1
Russian	1	1	1	1	1	1	1	7
Samoan				1				1
Tagalog		1	1	1	1		1	5
Thai				1				1
Ukrainian							1	1
Vietnam		1	1	1	1	1	1	6
Totals	7	9	15	20	13	9	18	

Table 1
Summary of "Federal Citizen Information Center" Multilingual Resources by Subject
http://www.pueblo.gsa.gov/multilanguage/multilang.htm?urlnet99

Governments have also been concerned with other types of multilingual activities. Some other emerging areas of government multilingual interest are terrorism and crime analysis. Terrorists use multiple languages, so to find them, systems need to consider and understand multiple languages. For example, Last et al. (2006) and Qin et al. (2006) investigated the use of multilingual approaches to discover the presence of terrorists groups on the Internet. Similarly, crime does not limit itself to a single language. Thus, Yang and Li (2007) discuss how to extract multilingual information for crime analysis focusing on Chinese and English documents.

3.4 Library Systems

Although they were not called knowledge management systems, perhaps the first real knowledge management systems were library systems. In particular, since knowledge management systems provide access to multiple resources, one comparable source is the library, although classic library-based research is seldom directly couched as knowledge management per se.

Extending commercial notions, including personalized portals such as "My Yahoo!," there has been a sequence of research from libraries that has been related to the development of personalized library portals. Starting with Morgan (1999) and Cohen et al. (2000) libraries have allowed users to create personal web pages to capture and store frequently used electronic library resources. There have been a number of updates to that original concept and views of the future (e.g., Ciccone 2005). In addition, there have been multilingual views of the "My Library" concept. For example, as noted by Rozic-Hristovski et al. (2002, p. 157), "One of the most important needs of visitors from … abroad is multilingual support, which means that the users can select a language in which the portal interface is presented to them."

3.5 Medical Systems

Sevinc (2005) and others have stressed the need for medical research to be available in multiple languages. Further, there have been some multilingual systems developed for support of medical problems. For example, Goble et al. (1994) created a multilingual terminology server designed to provide an ontology to a broad range of medical applications. As another example, Zhou, Ain and Chen (2006) focused on facilitating the search for Chinese medical information.

3.6 International Organizations

Some organizations are by their very nature "International." Those organizations also need to provide a range of multilingual knowledge resources. O'Leary (2007a) provides an in-depth case study of one such organization associated with the United Nations, Food and Agriculture Organization (FAO). That research provides a detailed analysis of many of the multilingual issues in a large international organization. Further, that analysis examines architecture and work flow issues associated with the implementation of multilingual systems in an extensible mark-up language (XML) structure. That research provides a benchmark and some detailed examples about multilingual systems.

4 Presentation of Multilingual Resources: Connect

In general, presentation has a connection function, connecting users with knowledge resources. There are a number of presentation issues associated with multilingual resources.

4.1 Languages and Content Availability

Although multilingual systems may provide access to resources in multiple languages, not all resources are necessarily provided in each language to the same extent. A summary of the key issues of this section is provided in figure 5.

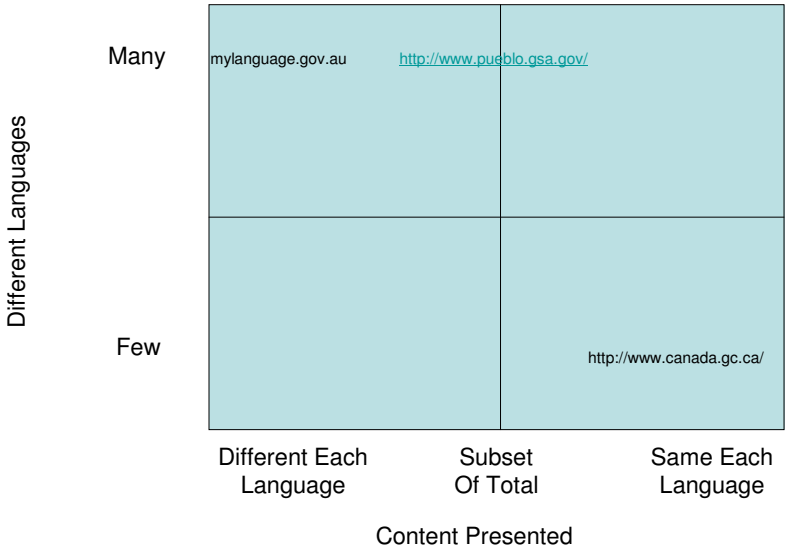

Figure 5
Presentation of Multilingual Resources

At some sites all of the knowledge resources are presented for each of the languages that are accommodated. For example, at the web site for the Canadian government (http://www.canada.gc.ca/) it appears that all of the knowledge resources are provided equally in both English and French, the two official languages of the country.

At other web sites only a subset of the total knowledge resources are provided for any one language (e.g., http://www.pueblo.gsa.gov/multilanguage/multilang.htm?urlnet99), as seen in table 1. As another example, at the Belgium web site (http://www.belgium.be/eportal/index.jsp) resources are given in four different languages (French, Dutch, German and English). However, not all resources are equally available for all languages. As a result, they provide the ability of a user to choose a "back-up" language. As noted on that site, "For the moment only a limited amount of content is available in English. In order to browse all the content of the federal portal, we suggest you to choose one of the other languages available below. Once you've done this, only the unavailable content in English will permanently be displayed in the language you have selected."

At still other sites the multilingual content in the different languages may vary substantially. For example, at mylanguage.gov.au, there is very limited overlap between the content available for the many different languages on the site.

4.2 Languages and Connection Links

A number of approaches have been used to capture multilingual links. As seen in figure 1, links to other language resources are oftentimes listed on the web page that presents only information in that language, as "English," etc., with the exception of an indicator as to language. Another approach seen in figure 6 puts the links of multiple languages side by side so the user can chose which is appropriate. This approach provides a notion that the resources are transparent, and equitable, in that neither language is provided with greater knowledge resources.

Figure 6
Slovenia Web Page (http://www.gov.si/)

Another approach is to provide links to foreign language content in another language. For example, in the following figure 7, links are to French content (http://www.pueblo.gsa.gov/multilanguage/multilang.htm#French). However, some of the links are in English and some are in French. It is not clear if English links to French materials are an effective means of presenting the material, and will generate use of foreign language content. Further, such issues also must be considered for search, which includes links and content.

Figure 7
US Links to French Content Materials

4.3 What Makes a Page?

Presentation often is thought of as translation. Translating different languages can result in documents of different lengths. Presentation then needs to assess if the different lengths are appropriate. Further, cultural issues need to be brought to the presentation of information to ensure that the resulting translation and presentation is appropriate for the setting.

5 Translation: Convert and Connect

Translation will help meet capabilities of conversion and connection. For example, resources in one language are translated to those of another language, and users in one language are connected to those resources, e.g., through collaboration or other approaches.

5.1 Machine Translation

Machine translation has been the source of substantial research (e.g., Nirenburg et al. 1994 and others). As a result, it is substantial enough for a survey of itself, without other knowledge management topics. Accordingly, it is beyond the scope of the current paper. However, machine translation provides hope that resources in one language can be rapidly and inexpensively converted to other languages. Further, through activities, such as collaboration, machine translation can help connect people to people.

5.2 Limitations of Multilingual Machine Translation

Unfortunately, there are a number of limitations associated with using translations for multilingual systems (e.g., Inaba 2007 and Caracciolo et al. 2007). First, translating different languages is not necessarily "symmetric." For example, the French word "banque" can be translated into English word "bank." However, the English word "bank" can refer to a place where money is kept or a place besides a river. As a result, in the later case, the appropriate translation from English to French would be "rive." Such errors and ambiguities are more likely to be introduced into translated text the more languages into which a language is translated. Second, translating different languages is not necessarily "transitive." For example, using Google, "River Bank" in English translates to "Rive" in French, which translates to "Ufer" in German, which translates to "Shore" in English. Third, there are differential translation capabilities for different languages. The extent of translation capabilities and the quality of those capabilities is not uniform across all language pairs. Accordingly, multilingual collaboration using translation may not result in the desired results. Fourth, acronyms do not translate well, since words in different languages are likely to start with different letters and/or occur in a different order. As a result, unless the acronym become accepted words in multiple languages, they are difficult to translate. Fifth, translations of official names as opposed to unofficial and shortened names also must be accounted for (e.g., United States, vs. United States of America).

6 Multilingual Collaboration: Content, Connecting and Converting

One approach to connecting users to users or users to content, and converting what they know to explicit knowledge is collaboration. Increasingly, collaboration attempts to facilitate multilingual

capabilities. However, at this time, multilingual collaboration is primarily occurring in the research labs.

Multilingual collaboration occurs in a number of forms. In its easiest to implement multilingual form, multilingual users contribute in multilingual settings. Unfortunately, the number of multilingual users is limited and the number of languages that any one individual communicates in is limited. As a result, in some cases users can employ their own language and that language will be translated.

6.1 Users Contribute in Multiple Languages

Perhaps the easiest from a system perspective is to facilitate collaboration where the users are multilingual. In this setting, users might be able to go from one conversation to another collaboration to another collaboration, independent of translation capabilities.

6.2 Users use their own language and it is translated

Unfortunately, generally, we cannot count on the user to have multilingual capabilities. Even if they have multilingual capabilities, they are likely to have access to a limited number of languages, and that access is likely to be stronger in some languages than others.

As a result, there has been substantial research examining how to facilitate multilingual collaboration through translation. Nomura et al. (2003) tested communication and collaboration in five different languages on a multilingual bulletin board system. They found that machine translations were problematic, impairing communication. As a result, they allowed multilingual users to modify translated sentences to improve the overall level of the translation. Funakoshi et al. (2003) developed a tool with which they were able to experiment with multilingual collaboration. They found that although translation may be appropriate for overall and high level discussions, that it was not appropriate for "detailed" discussions of cooperative works. Inaba et al. (2007) proposed a "language grid" structure to support multilingual content, with an active human user community that, like in Nomura et al (2003) allowed human participation in addition to the machine translation.

6.3 Human – Machine Integration

The limitations of machine translation led Nomura et al. (2003), Inaba et al. (2007) and others to describe a number of collaboration tools for multilingual environments. Throughout, their use of tools integrates human and machine capabilities to facilitate a better understanding of the necessary translation in an effort to improve multilingual collaboration.

7. Categorization and Definition: Content and Connecting

Knowledge categorization/definition, help facilitate content and help connect content to users. The primary contemporary approach to categorizing and defining knowledge is done using ontologies.

Ontologies and semantic devices, such as controlled vocabularies also facilitate multilingual knowledge management. In particular, each can be used to facilitate multilingual content and

multilingual search that can connect users to content and other users. Further, each can be used to help standardized content across multiple languages.

7.1 Ontologies

Ontologies are used to specify content and facilitate search. As a result, they are critical to multilingual environments. Gruber (1993) referred to ontologies as an explicit specification of a conceptualization. Gruber (1993) also suggested "contexts" to capture local views of a domain, as opposed to the global view of an ontology.

Kahng and McLeod (1996) realized that building a shared ontology would never be easy. As a result, they suggested creating ontologies with static and dynamic aspects. The static components would be mutually understood between participants, while the dynamic aspect evolves either by adding ontologies or discovering them. Using this approach, more than a single name can be given the same concept.

Segev and Gal (2007) based their work on Kahng and McLeod (1996) and Gruber (1993) when they proposed an ontology-based model of multilingual applications, with static components and dynamic components. The static components would be that portion mutually agreed upon, while dynamic would relate to a particular language. Their model was based on a global ontology that was manually designed for a specific domain. In addition, their model used local contexts, to further specify the ontology. The combination of ontologies and contexts lends itself to multilingual applications, where a single global ontology fails to capture all of the nuances that stem from language and cultural differences.

7.2 Example: WordNets

There are a number of available multilingual concept-based dictionaries available world wide. Fensel (2004) briefly discusses that a multilingual version of EuroWordNet (http://www.illc.uva.nl/EuroWordNet/). The word nets are structured nets of synonymous words with semantic relations between them. Because they are linked, it is possible to go from words in one language to other similar words in another language. Among the different word nets there is a static top shared ontology.

7.3 Example: Controlled Vocabularies such as AGROVOC

Lauser et al. (2002) examine some of the issues associated with developing an ontology used in a multilingual environment, with AGROVOC, illustrated in figure 8. As noted on the FAO web page (http://www.fao.org/aims/ag_intro.htm) "AGROVOC is a multilingual, structured and controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment)." As seen in the exhibit, apparently, some versions are FAO resources (e.g., the English, French, etc. versions), while others are controlled at a country-specific location, e.g., the Lao version.

According to Caracciolo et al. (2007), translations are provided by native speakers of the target language. Translations are typically made off of the English version and sent to FAO for validation and inclusion in the master version. Apparently, terms are assigned a unique number,

e.g., “Abalone” is assigned the number 5 in English. The translation of the word, e.g., in French “ormeau” is also provided the number 5 in the French version. As a result, the vocabulary is not alphabetically ordered in each language (see figure 9), but multiple names are attached to a single concept across the languages. AGROVOC is implemented using inheritance and a relational database structure.

Figure 8
AROVOC

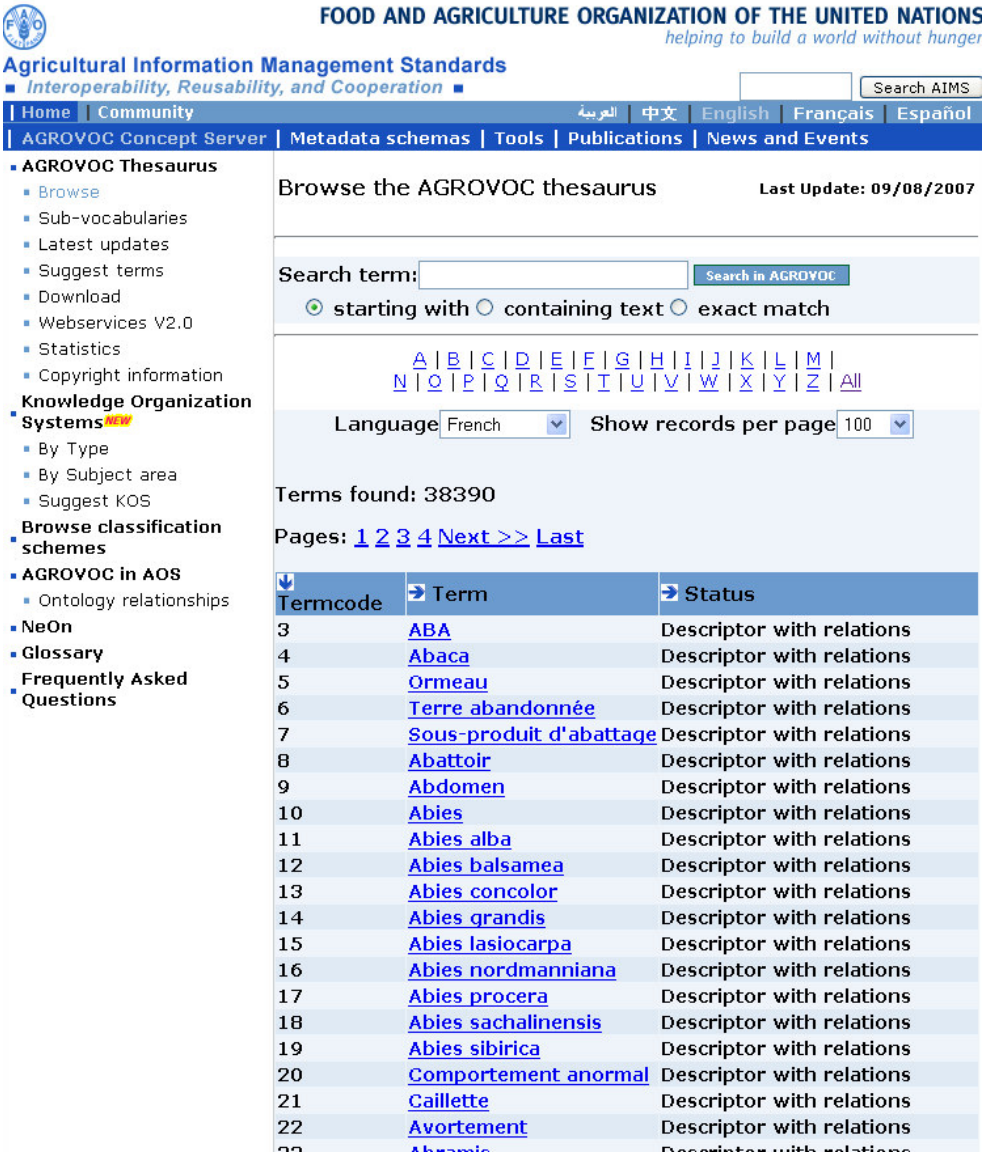

Figure 9
List of French Version of AGOVOC

However, the number of terms in the AGROVOC vocabulary apparently varies by language, with the number of entries for FAO controlled versions by language listed in the following table 2. Interestingly, the number of terms in the vocabulary varies substantially by language. The

differences can be the result of differences in language, but also control over additions to the vocabularies in those languages.

Language	November 1, 2007
Arabic	25948
Chinese	36794
Czech	39466
English	39613
French	38390
Japanese	38659
Portuguese	36347
Spanish	41714
Thai	25420

Table 2
AGROVOC Vocabulary Size in Different Languages

Information about the stability of such vocabularies is limited. Information about the impact of that stability on the use or expansion of these vocabularies also is limited. Further, based on these different sizes of the vocabularies, either some vocabularies are under-specified or some are over-specified, or characteristics of languages differentiate themselves from other language vocabularies for the same set of concepts.

8 Multilingual Search: Connecting and Content

Search connects the user to the knowledge content. Issues include, but are not limited by concerns over the number of letters in the alphabets being searched. Not surprisingly, the search literature is substantial and complete analysis is beyond the scope of this paper. Here we will examine some specific issues that differentiate multilingual search.

8.1 Selected Issues in Multilingual Search

There are a number of issues associated with multilingual search. First, is the user interface of the search engine available in multiple languages? As seen in figure 10, one search engine at FAO provides interfaces in five different languages.

Second, does the search engine allow the user to chose the language of materials that they want to search for? For example, does a search engine permit the use of an English interface, in the search for French materials about a particular topic? Third does the search engine provide a user with a list all of the relevant materials, independent of language. For example, when using a French language interface, should the default be to search for French language materials? These and other issues can be addressed in the analysis of multilingual search.

Figure 10
Search in an FAO Knowledge Base

8.2 Non-English Search

There has been substantial research on search in general, and search in multilingual settings (e.g., Savoy 2005). For example, Bar-Ilan and Guttman (2003) analyzed the ability of three different search engines to handle queries in four non-English languages. The found that content from languages that were not English had a larger chance of being lost in cyberspace.

Non English search is difficult for a number of reasons. As an illustration, there can be special characters in languages. For example, Aytac (2005) noted the difficulty of doing search in Turkish because of those special characters. As another illustration, Caracciolo et al. (2007) review some other issues, including the need to use particular character encoding capabilites (UTF-8 vs. UTF-16) and the need to support left to right and right to left languages.

8.3 Multilingual Portals

Portals provide users with a summary of key available knowledge resource materials. From a multilingual perspective, portals can be simply a multilingual interface structure to non multilingual content, or the portals can lead to multilingual content. An example of a multilingual portal interface is given in figure 11.

Multilingual portals can function as a multilingual user interface, putting all of the same content in different languages. On the other hand, multilingual portals also could be constructed by ensuring that the interface and the linked content are multilingual. A decision must also be made as to whether cultural differences are sufficiently large so that the information linked to and listed on the page should be the same for each language.

Figure 11
Portal with Multilingual Capabilities

9. Extensions and Research Issues

Using the research summarized in this paper there are a number of research issues that require additional research. Many of the additional research questions relate directly to the topics discussed above, presentation, translation, collaboration, categorization/definition and search including the following topics.

Presentation: What is the impact of delaying multilingual disclosures until all languages are ready? How effective is it to present links to foreign language material not in the foreign language? If resources are not available at the same time for multiple languages, under what conditions should they be made available to their users – should all versions be released at the same time?

Translation: How efficient and effective are human – machine based translation systems?

Collaboration: How can we overcome problems of translation in order to facilitate multilingual collaboration?

Categorization/Definition: What kinds of ontologies work best in a multilingual environment? How much change occurs in ontologies and does that inhibit ontology use? What portion of a multilingual ontology is static and what portion needs to be dynamic?

Search: How effective are multilingual portals in guiding users to multilingual resources? How effective is a multilingual interface that leads to a single language of knowledge resources? One key issue is should, and to what extent does search include multilingual content?

Further, we can use the theoretical framework to generate additional research issues, such as the following:

Content: Does the use of multiple languages provide an increase in transparency and trust?

Connection: Does including multilingual content as part of search increase transparency and trust?

Conversion: Does converting knowledge resources from one language to another, create transparency and trust?

10 Summary and Contributions

This paper has summarized some of the primary multilingual knowledge management literature around a basic model of knowledge management capabilities of content, connecting and converting (e.g., O'Leary 1998a). Further, this paper has extended some of the theory of knowledge management beyond that of early developments to account for multilingual aspects of knowledge management. In addition, this paper investigated how those capabilities were implemented in some of the key knowledge management activities, including presentation, translation, collaboration, categorization/definition and search. Applications in a number of domains were summarized. In addition, this paper has pointed to a number of potential research topics.

References

Bar-Ilan, J. and Gutman, T., "How do search engines handle non English Queries? A Case Study," in Proceedings of the Twelfth International World Wide Web Conference, 2003.

Caracciolo, C., Sini, M., Johannes Keizer, J., "Requirements for the treatment of multilinguality in ontologies within FAO," ftp://ftp.fao.org/docrep/fao/010/ah894e/ah894e.pdf, October 2007.

Chau, R., Yeh, C-H., Smith-Miles, K., "Fuzzy-Neuro Web-based Multilingual Knowledge Management," in L. Wang et al. 3rd International Conference on Fuzzy Systems and Knowledge Discovery - FSKD, Lecture Notes in AI, 4223, pp. 1229-1238, Springer – Verlag, Berlin Heidelberg, 2006.

Ciccone, K., "MyLibrary @ NCState,: A Library Portal After five Years," Journal of Library Administration, Volume 43, Numbers 1 / 2, pp. 19-35, 2005.

Cohen, S., Fereira, J., Horne, A. Kibbee, B. Mistlebauer, H. and Smith, A., "MyLibrary: Personalized Electronic Services in the Cornell University Library," D-Lib Magazine, April 2000, Volume 6 Number 4, http://www.dlib.org/dlib/april00/mistlebauer/04mistlebauer.html

Fensel, D., Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer, Berlin, Second Edition, 2004.

Funakoshi, K., Yamamoto, A., Nomura, S., and Ishida, T., "Lessons Learned from Multilingual Collaboration in Global Virtual Teams," http://www.ai.soc.i.kyoto-u.ac.jp/publications/03/kaname-hcii2003.pdf, Tenth International Conference on Human Computer Interaction, 2003.

Goble, C., Crowther, P. and Solomon, D. " A Medical Terminology Server," in Database and Expert Systems Applications, Springer Verlag Volume 856, 1994.

Gruber, T., "A translational approach to portable ontologies," Knowledge Acquisition Volume 5, Number 2, 1993, pp. 199–220.

Inaba, R., Murakami, Y., Nadamoto, A., and Ishida, T., "Multilingual Communication Support Using the Language Grid," in T. Ishida, S. Fussel, and P. Vossen (Eds): IWIC 2007, Lecture Notes in Computer Science, 4568, pp. 118-132, 2007, Springer – Verlag, Berlin.

Kahng, J. and McLeod, D., "Dynamic classification ontologies for discovery in cooperative federated databased, in Proceedings of the First International Conference on Cooperative Information Systems, June 1996, pp. 26-35, Brussels, Belgium.

Last, M., Markov, A. Kandel, A., "Multi-lingual Detection of Terrorist Content on the Web," in H. Chen et al. (Eds.) WISI, Lecture Notes in Computer Science, 3917, 2006, pp. 16-30, Springer – Verlag.

Lauser, B. Wildeman, T., Poulos, A., Fisseha, F., Keizer, J., Katz, S., "A Comprehensive Framework for Building Multilingual Domain Ontologies: Creating a Prototype Biosecurity Ontology," Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002, pp. 113-123, Firenze University Press.

Morgan, E. "MyLibrary @ NCState," Proceedings of the Customized Information Delivery Workshop, SIGIR, Berkeley, CA August 19, 1999, pp. 12-18, http://infomotions.com/musings/sigir-99/

Nirenburg, S., Carbonell, J., and Tomita, M., Machine Translation: A Knowledge-based Approach, Morgan Kaufman Publishers, San Francisco, USA, 1994.

Nomura, S., Ishida, T., Yosuyoka, M., Yamashita, N., Funakosh, K., "Open Source Software Development with your Mother Language: Intercultural Collaboration Experiment 2002," in Proceedings of the International Conference on Computer Supported Cooperative Work, HCI2003, pp. 1163-1167, 2003.

O'Leary, D. E. "Knowledge Management Systems: Converting and Connecting," IEEE Intelligent Systems, May – June 1998a, pp. 30-33.

O'Leary, D.E., "Using AI in Knowledge Management: Knowledge Bases and Ontologies," IEEE Intelligent Systems, May – June 1998b, pp. 34 – 39.

O'Leary, D. E., Enterprise Resource Planning Systems, Cambridge University Press, 2000.

O'Leary, D. E. "A Multilingual Knowledge Management System: A Case Study of FAO and WAICENT," Decision Support Systems, Forthcoming, 2007a (available on line July 27, 2007).

O'Leary, D. E., "Evolution of Knowledge Management Toward Enterprise Decision Support: The Case of KPMG," F. Bernsein and C. Holsapple, Editors, Handbook on Decision Support Systems, 2007b, pp. 1135-1162.

O'Leary, D. E., "Multilingual E-Government," Submitted for publication, September, 2007c.

O'Leary, D. E., "Multilingual Web Presence: Case of the Fortune 25," Forthcoming, WeB 2007, Montreal, December 2007.

Otterbacher, J., "Adoption of Translation Support Technologies in a Multilingual Work Environment," In T. Ishida, S. R. Fussell and PTJM Vossen, (Eds), IWIC 2007, Lecture Notes in Computer Science, 4568, pp. 276-290, Springer – Verlag, Berlin Heidelberg 2007.

Peters, C., and Sheridan, P., "Multilingual Information Access," in M. Agosti, F. Crestani and G. Pasi (Eds.), ESSIR 2000, Lecture Notes in Computer Science 1980, pp. 51-80, 2000, Springer – Verlag.

Qin, J., Zhou, Y., Reid, E. Lai, C. and Chen, H., "Unraveling International Terrorist Groups Exploitation of the Web," in H. Chen et al. (Eds.) WISI, Lecture Notes in Computer Science, 3917, 2006, pp. 4-15, Springer – Verlag.

Rozic-Hristovski, A., Humar I., and Hristovski, D., "Developing a Multilingual, Personalized Medical Library Portal: Use of MyLibrary in Slovenia," Electronic Library and Information Systems, Volume 37, No. 3, pp. 146-157, 2002.

Savoy, J., "Comparative Study of Monolingual and Multilingual Search Models for Use with Asian Languages," ACM Transactions on Asian Language Information Processing, Vol. 4, No. 2, June 2005. Pages 163-189

Segev, A., and Gall, A., "Enhancing Portability with Multilingual Ontology-based Knowledge Management," Decision Support Systems, Forthcoming, 2007 (Available on line August 6, 2007).

Sevinc, A., "Multilingual Approach to 'Web of Science,'" Journal of the National Medical Association, Volume 97, No. 1, January 2005, pp. 116-117.

Yang, C. C., Wei, C-P, Li, K.W., "Cross Lingual Thesaurus for Multilingual Knowledge Management," Decision Support Systems, Forthcoming, 2007.

Yang, C. and Li, K., "An Associate Constraint Network Approach to Extract Multi-lingual Information for Crime Analysis," Decision Support Systems, Volume 43, Issue 4, August 2007, Pages 1348-1361.

Zhou, Y., Qin, J. and Chen, H., "CMedPort: An Integrated Approach to Facilitating Chinese Medical Information Seeking," Decision Support Systems, Volume 42, Issue 3, December 2006, Pages 1431-1448.

Zisman, A., Chelsom, J., Dinsey, N., Katz, S., Servan, F., "Using Web Services to Interoperate Data at FAO," Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002, pp. 147-156. http://www.bncf.net/dc2002/program/ft/paper17.pdf