

*“Usted dice “palta”, yo digo “aguacate”
y ellos dicen “avocado”*

Diversidad en la terminología agrícola de las Américas*

Lori Finch¹ y Melanie Gardner²

Resumen

Un tesoro es un instrumento de control de palabras que se utiliza para organizar términos y expresar relaciones entre conceptos. Durante décadas, especialistas de la información han dependido de los tesauros para contribuir a estandarizar la terminología en sistemas de recuperación de datos. En este artículo se discute la importancia y la necesidad de realizar un esfuerzo conjunto para elaborar un tesoro y glosario inglés/español que refleje las variaciones locales del lenguaje utilizadas en los países de ALC en materia agrícola. En el 2006, la Biblioteca Agrícola Nacional de los Estados Unidos (NAL) y la Biblioteca Conmemorativa Orton (BCO) del IICA empezaron a trabajar conjuntamente en ese esfuerzo y, en mayo del 2007, publicaron un tesoro bilingüe. Hasta la fecha, dichas instituciones han lanzado un WIKI, han identificado un flujo de trabajo inicial y están aprendiendo a trabajar a través de las distancias y los husos horarios para crear una herramienta que aumente el acceso a la información agrícola a lo largo y ancho de las Américas.

1 Coordinadora del Thesaurus, Biblioteca Nacional de Agricultura del Departamento de Agricultura de los Estados Unidos. lori.finch@ars.usda.gov

2 Coordinadora de AgNIC, Biblioteca Nacional de Agricultura del Departamento de Agricultura de los Estados Unidos, melanie.gardner@ars.usda.gov

* Se agradecen los aportes de Priscila Cascante del IICA.

Palabras clave: *Tesoros, terminología, agricultura, América Latina, Caribe.*

Introducción

Científicos e investigadores han analizado las implicaciones de permitir que los datos y los resultados de sus investigaciones estén disponibles de forma gratuita y públicamente mediante el servicio de acceso abierto. La viabilidad y disponibilidad de la información depende del ambiente social, económico, político, contextual y cultural dominante. Sin embargo, la cantidad de información científica disponible actualmente de forma gratuita por medio del acceso abierto ha aumentado sustancialmente en la última década. A ello han contribuido los eruditos de América Latina y el Caribe (ALC), tal como señalan Alperin, Fischman y Willinsky (2008).

Los científicos están dando grandes pasos en la publicación de datos y resultados científicos a través de infraestructura de las tecnologías de la información y la comunicación (TIC). Actualmente, la investigación agrícola que se ofrece a través del acceso abierto se disemina fácilmente entre quienes tratan de solucionar problemas reales en este campo, tales como el diagnóstico de enfermedades de las plantas y métodos de control.

Sin embargo, no es suficiente contar con mayor información mediante herramientas TIC, si se dificulta la búsqueda de determinados datos. Dado que se dispone de una mayor cantidad de información en la *web*, será más difícil para el investigador, el educador, el estudiante o el científico

Actualmente, la investigación agrícola que se ofrece a través del acceso abierto se disemina fácilmente entre quienes tratan de solucionar problemas reales en este campo, tales como el diagnóstico de enfermedades de las plantas y métodos de control.



encontrar los datos requeridos si no se cuenta con las herramientas necesarias para la búsqueda. Aunque el servicio de acceso abierto y la tecnología para enviar

datos a teléfonos celulares constituyen un gran esfuerzo para diseminar información científica a quienes se encuentran conectados al sistema, ello no soluciona el problema de la complejidad del idioma ni de la necesidad de lograr de manera eficaz la recuperación de datos.

Complejidad del idioma

La complejidad del idioma debido a las múltiples variedades del lenguaje en el español de América constituye un escenario multilingüístico que debe ser considerado en la configuración de tesauros agrícolas, como se observa en este simple ejemplo:

Cliente: —*Hola, estoy buscando artículos sobre enfermedades fungosas de la palta.*

Bibliotecario: —*¿Palta?*

Cliente: —*Persea americana.*

Bibliotecario: —*¡Oh, sí! ¡aguacate! enfermedades fungosas del aguacate. Ahora entiendo.*

Por lo tanto, no es de extrañarse que quien cultiva aguacate en Perú o Chile y usa la palabra “palta” para referirse a la fruta de la *Persea americana*, no reconoce el uso del término “aguacate” que utilizan los centroamericanos. Es común que los hablantes del español como lengua

materna, provenientes de distintas regiones, no comprendan las palabras que se utilizan fuera de su propia región, debido a las mezclas lingüísticas de sus hablantes.

Desde este punto de vista, la información agrícola proveniente de las diversas formas lingüísticas utilizadas por los hablantes de las Américas, responde también a su cultura, a sus necesidades de comercialización y a los mecanismos de supervivencia.

Si se considera que los idiomas más importantes que se hablan en las Américas son el inglés y el español y que este último es el idioma de mayor crecimiento utilizado en Estados Unidos en el campo de la agricultura, se hace necesario contar con herramienta bilingüe estandarizada que le permita al usuario utilizar de manera eficaz la información para el enriquecimiento e intercambio de información en las Américas.

El IICA y la Biblioteca Nacional de Agricultura (NAL) han reconocido esta necesidad, por lo que se han asociado para desarrollar una herramienta bilingüe en la que se incorpore la mayor cantidad de variedades del lenguaje del español de América. Este recurso, conocido como Tesoro Agrícola, ha estado disponible en línea desde mayo del 2007 y contiene más de 70 000 términos sobre agricultura y disciplinas complementarias.

 *La información agrícola proveniente de las diversas formas lingüísticas utilizadas por los hablantes de las Américas, responde también a su cultura, a sus necesidades de comercialización y a los mecanismos de supervivencia.*

Información del Tesauro Agrícola de la NAL

- Bilingüe, español/inglés.
- Énfasis en términos locales de ALC.
- Más de 70 000 términos.
- Nombres científicos y comunes de organismos.
- Definiciones de términos técnicos disponibles en un glosario adicional.
- Cobertura detallada de conceptos agrícolas y biológicos.
- Características especiales de indexadores, como los números de clasificación de enzimas y los códigos del Comité Internacional de Taxonomía de los Virus (ICTV).
- Colaboración del IICA, la BCO y el USDA.
- Descarga gratuita de archivos en formatos XML, PDF, SKOS y MARC en el sitio web http://agclass.nal.usda.gov/agt_es.shtml
- Actualizado anualmente desde el 2002 en el mes de enero.
- Disponible en internet las 24 horas al día, los siete días de la semana, con un sitio espejo de respaldo en la Universidad Estatal de Michigan.

El tesauro es una herramienta de control de palabras que se utiliza para organizar términos y expresar relaciones entre conceptos. Durante décadas, especialistas en la información han dependido de los tesauros para contribuir a estandarizar la terminología en sistemas de recuperación de datos (Gilchrist, Lancaster, Lancaster y Warner).

Además, el tesauro constituye un vocabulario controlado, en el que cada término representa un concepto. Sirve de lenguaje para la indización de un sistema de recuperación de datos, en el que se encuentra la colección de términos utilizados para expresar el contenido temático de los ítems léxicos en dicho sistema. Por ejemplo:

Título del artículo: Control de enfermedades fungosas en cultivares de aguacate en el Valle de Chanchamayo

Términos temáticos del vocabulario controlado que describen el contenido temático del artículo:

Tema: *Persea americana*

Tema: aguacate

Tema: enfermedades fungosas

Tema: cultivar

Tema: Perú

Tema: control de enfermedades de las plantas

El tesauro es una herramienta de control de palabras que se utiliza para organizar términos y expresar relaciones entre conceptos.



► **Los sistemas de recuperación de datos que utilizan un vocabulario controlado estándar como el tesoro facilitarán al usuario la tarea de encontrar datos.**

El proceso para asignar términos temáticos a los ítems se denomina indización. La indización de temas agrega valor a un sistema de recuperación de datos, de manera que los ítems se pueden encontrar más fácilmente.

La terminología que se maneja en el campo de la agricultura responde a una jerga especializada que, por lo general, es manipulada de manera diferente por los hablantes comunes. En algunas jergas técnicas, como en la nomenclatura biológica, existen listas reconocidas por autoridades de nombres científicos válidos. Dicha estandarización es útil para la comunicación, como se mostró en el ejemplo anterior. El bibliotecario comprendió el término *Persea americana*, que constituye el nombre científico estandarizado. En dicho ejemplo, el concepto de “enfermedades fungosas” también se puede expresar de varias formas:

- Enfermedades fungosas
- Enfermedades micóticas de plantas
- Enfermedades por hongos

Para realizar búsquedas más eficaces de documentos de investigación sobre este tema, el usuario necesitaría usar las tres frases, lo cual le permitirá encontrar toda

la información pertinente relacionada con dicho tema. Los sistemas de recuperación de datos que utilizan un vocabulario controlado estándar como el tesoro facilitarán al usuario la tarea de encontrar datos, ya que una frase será utilizada consistentemente para el concepto de enfermedades fungosas.

El tesoro ofrece más que la simple estandarización de los términos, dado que agrupa los “términos similares”. Por ejemplo, se podrán encontrar agrupadas las palabras alcachofas, brócoli, coliflor, pepinos, cebollas y tomates en la lista relativa a “verduras” (Figura 1). Quien busca información sobre verduras no necesita recordarlas todas, ya que el tesoro proporciona una lista de ellas.

El tesoro también reúne sinónimos. Palta y aguacate son ejemplos de equivalentes regionales del mismo concepto. En un vocabulario controlado, un término constituye el término recomendado, es decir, el que se selecciona para describir el tema de los ítems en un sistema de información. En “palta” use “aguacate”, se les está indicando a los responsables de describir los ítems que utilicen el término aguacate.

Otro tipo de sinonimia que se maneja en el tesoro corresponde a las variantes ortográficas, lo cual se muestra en el siguiente ejemplo: *turfgrasses* o *turf grasses* (Figura 2). Dichas variantes son comunes en el idioma inglés y particularmente frecuentes entre el inglés estadounidense y el británico, como en *oestrogens* y *estrogens*. En el tesoro se escoge uno de ellos como término recomendado y se designa el otro como referencia cruzada.

Buscar el término	0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z	Categorías
<input type="text"/> <input type="button" value="Buscar"/>	<p>verduras</p> <p>Definición Cualquier parte de una planta que es ingerida comúnmente por los humanos como alimento, pero que no es considerada culinariamente como fruta, nuez, hierba, especia o grano.</p> <p>English vegetables</p> <p>Términos Específicos aceitunas aguacate alcachofas alubias verdes apio berenjenas brócoli brotes de bambú brotes de granos calabazas cardo cebollas coles de Bruselas coliflor fruta de pan fruta del pobre gajos de espárragos gombo hinojo hongos comestibles maíz dextrinoso pepinos pimientos puerros ruibarbo tomates tubérculos comestibles vegetales de hoja verde zapallos</p> <p>Términos Genéricos productos a base de hortalizas</p> <p>Términos Relacionados cultivos vegetales jardines de verduras jugos de hortalizas</p>	<p>Cambiar de Display</p> <p>Mostrar la Jerarquía . = Términos Específicos : = Términos Genéricos</p> <p>Busque su término</p> <p>AGRICOLA Artículos AGRICOLA Libros Google Académico</p>
<p>Opciones de la búsqueda</p> <p>Lengua español / Spanish</p> <p>Método de búsqueda Términos que incluyen..</p> <p>Número de términos desplegados 100</p>		

Fig. 1. Extracto del Tesoro Agrícola, en el que se muestran los términos similares agrupados en jerarquías.

The screenshot shows the website interface for the 'tesauro' (thesaurus) provided by the USDA National Agricultural Library and the Inter-American Institute for Cooperation on Agriculture (IICA). The page is titled 'Search Results' for the term 'turf grasses'. On the left, there is a search bar with the text 'Search for Term' and a 'Search' button. Below the search bar are 'Search options' including a 'Language' dropdown set to 'English', a 'Search type' dropdown set to 'terms contain text', and a 'Terms per page' dropdown set to '100'. The main content area displays the search results for 'turf grasses', including a 'Spanish' entry 'gramíneas para céspedes', a 'Used For' entry 'turfgrasses', a 'Broader Term' entry 'grasses', and a 'Related Term' entry 'lawns and turf thatch turf management'. On the right side, there is a 'Change Display:' section with options for 'Show Hierarchy' (Narrower Terms and Broader Terms) and a 'Search your Term in:' section with links to 'Google Scholar', 'AGRICOLA Articles', and 'AGRICOLA Books'. The footer contains various links including 'NAL | IICA | USDA | Agricultural Research Service | Science.gov | GPO Access | Web Policies and Important Links | Site Map | FOIA | Accessibility Statement | Privacy Policy | Non-Discrimination Statement | Information Quality | FirstGov | White House'.

Fig. 2. Extracto del Tesauro Agrícola, en el que se muestran las variantes ortográficas en inglés.

El tesauro no es un documento estático, sino más bien un recurso dinámico que debe mantenerse al paso de los descubrimientos en la agricultura y a los progresos técnicos. La NAL y el IICA, a través del personal de la Biblioteca Conmemorativa Orton (BCO), están colaborando para ampliar el vocabulario a fin de satisfacer todas las necesidades de América Latina. Sin embargo, se necesitan con urgencia expertos de ALC que contribuyan a ampliar este recurso mediante la inclusión de términos de su dialecto regional, de forma que su conocimiento pueda ser compartido.

Es fundamental que todos los países brinden aportes, de manera que otros

puedan aprovechar sus investigaciones para el beneficio común.

El IICA y la NAL reconocen la necesidad de trabajar conjuntamente, con el propósito de que esta herramienta léxica sea útil para lograr una mejor indización de la literatura agrícola y una mejor recuperación de datos sobre la agricultura de zonas tropicales y templadas para el uso de los agricultores de las Américas.

No solo se requiere agregar información al tesauro. También es necesario que expertos en el idioma y en el tema de la agricultura revisen la información existente. Dado que el idioma contiene homógrafos, un traductor podría entender mal el concepto que tiene a mano y traducirlo incorrectamente. Por ejemplo, el término inglés *bits* no brinda su significado a simple vista; el traductor necesitaría consultar la jerarquía y otras notas asociadas al término para poder traducirlo correctamente.

► *El tesauro no es un documento estático, sino más bien un recurso dinámico que debe mantenerse al paso de los descubrimientos en la agricultura y a los progresos técnicos.*

Un grupo proveniente de Chile completó la traducción original del Tesauro Agrícola. El traductor proporciona vocablos de uso común en esa región, pero no representa todas las variedades regionales de ALC. Se requiere de expertos provenientes de diversos países de ALC que revisen la terminología, de manera que haya una representación equitativa y completa de los dialectos regionales.

En el 2008, el IICA y la NAL comenzaron a utilizar un WIKI colaborativo, a fin de facilitar el desarrollo y el mantenimiento del tesauro. El WIKI es una nueva herramienta de la generación Web 2.0, que permite construir socialmente el conocimiento en temas de interés común. Constituye el espacio de retroalimentación de los “expertos” en lo relativo a terminología agropecuaria, utilizado para revisar y actualizar permanentemente el tesauro agropecuario producido por la NAL. Esta herramienta funciona como una “pizarra”, en la que los participantes pueden proponer nuevos términos y corregir errores en el tesauro, sugerir definiciones de términos y colocar traducciones. Las propuestas que se discuten sobre el WIKI serán incluidas en la edición 2010 del tesauro. Durante su primer año de uso, el WIKI ha sido útil para establecer flujos de trabajo, brindar mantenimiento al tesauro y permitir la participación de otros expertos agrícolas de ALC.

Dar mantenimiento a un tesauro implica la colaboración de expertos agrícolas en la materia así como especialistas en lexicografía; no obstante, quienes participen en este proceso también requieren aprender sobre los principios de la construcción de tesauros. La Organización Internacional de



Dar mantenimiento a un tesauro implica la colaboración de expertos agrícolas en la materia así como especialistas en lexicografía; no obstante, quienes participen en este proceso también requieren aprender sobre los principios de la construcción de tesauros.

Estandarización (ISO) es la encargada de establecer las directrices generales relativas a dicha construcción.

Por su parte, la NAL, con la ayuda de sus socios del IICA, se encarga de establecer las directrices y los principios específicos para construir el Tesauro Agrícola. Para ello se deben establecer principios y reglas para que haya consistencia en todo el tesauro. Por ejemplo, en abreviaturas, acrónimos, símbolos, puntuación, uso de las mayúsculas, nombres científicos,

nombres comunes, términos geográficos, forma de los términos y la desambigüedad de los homógrafos.

Por otra parte, el proceso de selección de términos es crítico y debe seguir una norma establecida. Los términos deben representar conceptos que tengan una buena aceptación en la disciplina, para lo cual se puede hallar evidencia en la literatura existente, por ejemplo, en *Agri2000*, *AGRÍCOLA* o *Google Scholar*, y determinar la frecuencia de su uso.

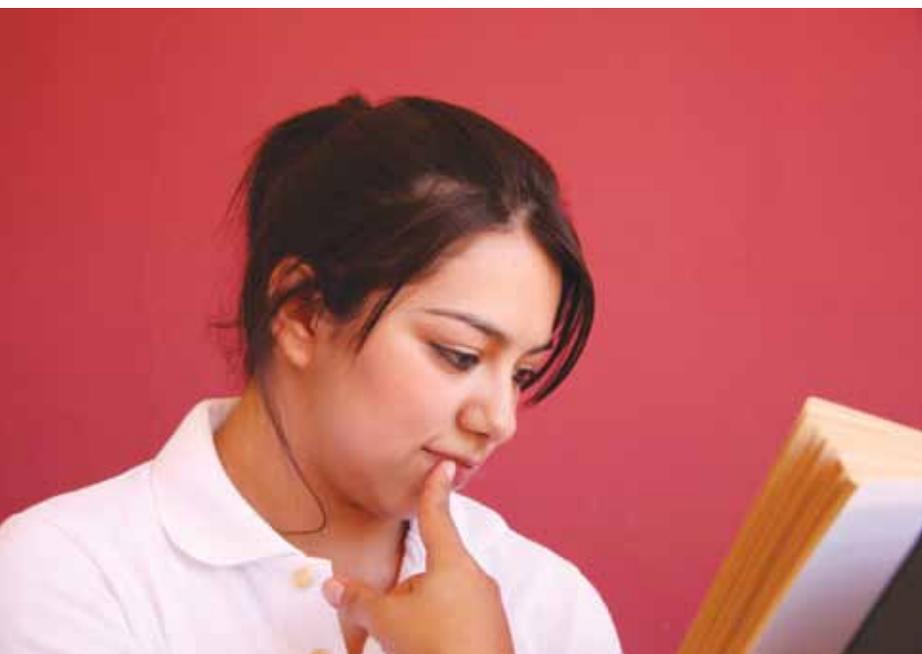
Otro método para encontrar términos apropiados consiste en examinar registros de búsqueda de los usuarios, llamado justificación por usuario existente. Se puede justificar la adición de un término al tesoro, si los usuarios lo buscan con frecuencia en un sistema de información. Los indexadores o quienes aplican el vocabulario controlado constituyen excelentes fuentes de nuevos términos.

Conforme aplican el vocabulario controlado, los indexadores pueden ver los nuevos términos y conceptos que se están generando en una disciplina. Es aconsejable consultarlos con ellos o analizar los términos no controlados de la materia asignados por los indexadores, con el fin de hallar términos candidatos para ser incluidos en el tesoro. No se debe subestimar la importancia de conocer las reglas establecidas para construir tesauros coherentes en su estructura y contenido.

Desafíos

La NAL y el IICA dependerán de los conocimientos especializados y del liderazgo de su personal para ampliar el proyecto, a fin de incluir una gran variedad de expertos provenientes de ALC. Actualmente, la BCO del IICA y la NAL trabajan para encontrar las mejores maneras de procesar eficientemente las propuestas de nuevos términos y cambios en el vocabulario. También evalúan tecnologías como la WIKI para comprobar si son útiles y apropiadas para realizar un proyecto con una gran cobertura geográfica. Es posible que se requieran otras tecnologías, especialmente para capacitar a los participantes sobre los aspectos prácticos del mantenimiento de los tesauros. Asimismo, la NAL y la BCO analizan las áreas del tesoro que requieren un mayor desarrollo y están identificando expertos que contribuyan a obtener resultados favorables.

El desafío y la oportunidad de una futura expansión y desarrollo del Tesoro Agrícola son inminentes. Los recursos humanos y el intelecto necesarios para realizar este trabajo no se encuentran



en una organización ni en un país de ALC en particular. Colectivamente, abunda el conocimiento, el talento y los conocimientos agrícolas especializados en las Américas. Se espera que el interés por alcanzar el éxito en este proyecto, que puede beneficiar a muchos, sea tan fuerte como el compromiso manifestado ya por algunos individuos y organizaciones en ALC.

También se espera concientizar al lector sobre la verdadera complejidad del lenguaje y la atención que merece. Ello motiva a mejorarlo y convertirlo en una herramienta que sea de utilidad para los sistemas de información agrícola de las Américas. Esta herramienta permitirá la comprensión sin importar si el agricultor que cultiva aguacate es de Perú, Guatemala o de Estados Unidos.

Literatura consultada

Ager, S. s.f. *Latin American Spanish or Spanish for Latin America*. Omniglot. Disponible en http://www.omniglot.com/language/articles/latin_american_spanish.htm

Alperin, JP; Fischman, GE; Willinsky J. 2008. *Open Access and Scholarly Publishing in Latin America: Ten flavours and a few reflections*. *Liinc em Revista* 4(2): 172-185. Disponible en <http://www.ibict.br/liinc>.

Chan, LM; Richmond, PA; Svenonius, E. 1985. *Theory of Subject Analysis: A Sourcebook*. Littleton, Colorado, US, Libraries Unlimited, Inc. 415 p.

Chamis Yanosko, A. 1991. *Vocabulary Control and Search Strategies in Online Searching*. *New Directions in Information Management* n.º 27:121. Nueva York, US, Greenwood Press.

Gilchrist, A. 1971. *The Thesaurus in Retrieval*. Londres, Aslib, 183 p.

International Standards Organization. *Documentation - Guidelines for the establishment and development of monolingual thesauri*. ISO 2788-1986. 32 p.

_____. *Documentation – Guidelines for the establishment and development of multilingual thesauri*. 61 p.

Lancaster, FW. 1986. *Vocabulary Control for Information Retrieval*. Arlington, Virginia, US, Information Resources Press. 270 p.

_____; Warner AJ. 1993. *Information Retrieval Today*. Information Resources Press, 341 p.

_____. 1998. *Indexing and Abstracting in Theory and Practice*. Champaign, Illinois, US, Universidad de Illinois. 412 p.

Onsrud, H. 2004. *Overview of Open-Access and Public-Commons Initiatives in the United States*. *Proceedings of an International Symposium on Open Access and the Public Domain in Digital Data and Information for Science* hosted by the Board on International Scientific Organizations.

Suber, P. 2007. *Open Access Overview: Focusing on open access to peer-reviewed research articles and their preprints*. Disponible en <http://www.earlham.edu/~peters/fos/overview.htm>

Wellisch, HH. 1991. *Indexing from A to Z*. Bronx, Nueva York: HW. Wilson Company, 461p.

Résumé / Resumo / Abstract

► « Pour désigner l'avocat, vous dites 'palta', je dis 'aguacate' et d'autres, 'avocado' » : La diversité dans la terminologie agricole des Amériques

Un thésaurus est un répertoire structuré qui sert à organiser des termes et à établir des relations entre des notions. Depuis des décennies, les spécialistes de l'information s'appuient sur des *thesauri* pour normaliser la terminologie des systèmes d'extraction d'information. Dans le présent article, nous examinons les fondements et la nécessité du nouveau partenariat qui s'est donné pour tâche d'établir un thésaurus et un glossaire anglais/espagnol rendant compte des variantes locales utilisées dans les pays d'Amérique latine et des Caraïbes. La *National Agricultural Library* (NAL), l'Institut interaméricain de coopération pour l'agriculture (IICA) et la Bibliothèque commémorative Orton (BCO) ont commencé à travailler de concert dans ce but en 2006. Un thésaurus bilingue a été publié en mai 2007. À ce jour, les partenaires ont lancé un WIKI et défini un ordonnancement des tâches, et ils apprennent à travailler à travers les distances et les fuseaux horaires pour créer un outil qui facilite l'accès à l'information agricole dans toutes les Amériques.

► Você diz palta, eu digo abacate e eles dizem avocado: Diversidade na terminologia agrícola das Américas

Tesouro é um vocabulário estruturado que serve para organizar termos e expressar relações entre conceitos. Especialistas em informação há décadas têm recorrido aos tesouros para ajudá-los na padronização da terminologia nos sistemas de recuperação da informação. Este artigo discute as bases e a necessidade desse novo esforço conjunto para desenvolver um tesouro e um glossário inglês/espanhol que reflita as variações locais representadas em toda a América Latina e o Caribe. A *National Agricultural Library* (NAL) dos Estados Unidos, o Instituto Interamericano de Cooperação para a Agricultura (IICA) e a Biblioteca Memorial Orton (BCO) iniciaram em 2006 uma parceria nesse sentido, havendo sido lançado um tesouro bilíngue em maio de 2007. Recentemente, esses parceiros lançaram um WIKI, identificaram um programa de trabalho inicial e estão aprendendo a lidar com as distâncias e os fusos horários para criar uma ferramenta que intensifique o acesso à informação em agricultura nas Américas.

► You say "palta," I say "aguacate" and they say "avocado": Diversity in agricultural terminology of the Americas

A thesaurus is an instrument for controlling words which serves to organize terms and to express relationships among concepts. For decades, information specialists have relied on thesauri to help with the standardization of terminology in information retrieval systems. This article discusses the importance of and need for a joint effort to develop an English/Spanish thesaurus and glossary which reflects the local variations in language found throughout the Americas in the area of agriculture. In 2006, the National Agricultural Library (NAL) of the United States and the Orton Commemorative Library (OCL) of the Inter-American Institute for Cooperation on Agriculture (IICA) began working together on this effort, and in May 2007 published a bilingual thesaurus. To date, the partners have launched a WIKI, identified an initial workflow and are learning how to work across distances and time zones to create a tool which enhances access to agricultural information across the Americas.