

# SciELO Machine Translation Integration

Helvecio Ribeiro – MSR-MT

Microsoft®  
**Translator**

# Who are we?

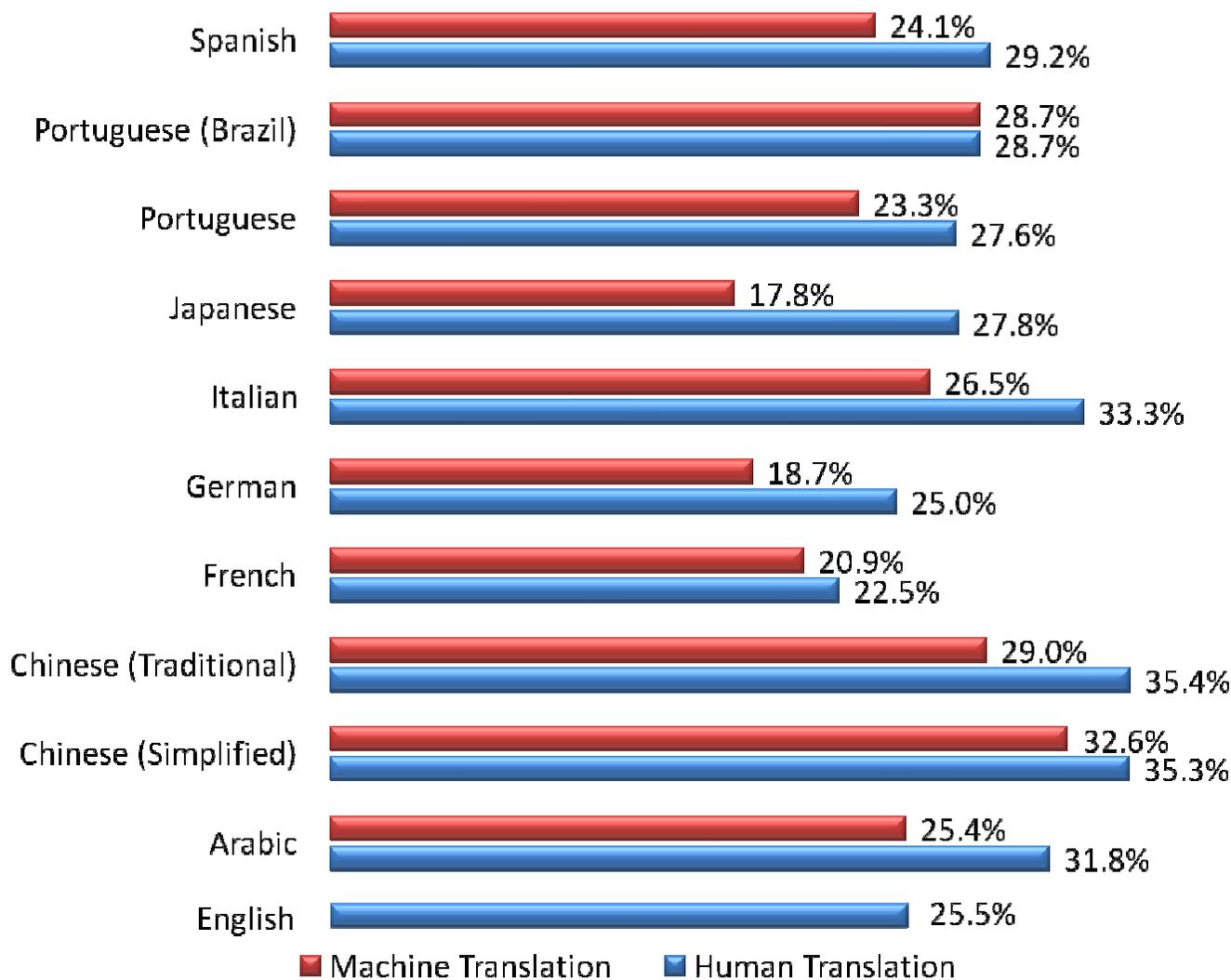
- MSR-MT is an incubation team part of the Natural Language Processing research group at Microsoft
- MT is a 10 years old initiative at MSR
- First active MT engagement was related to translating Knowledge Base articles
- Last year we released our first products open to the general public @ <http://translator.live.com>

# Time Line



# Now seriously...Why MT???

KB Survey Average resolve rate MT vs. HT



MT Quality Rating 1 - 4 (From English to Other Language)
ZH-TW: 2.81
ZH-CN: 2.82
PT-BR: 3.06
PT: 3.25
JA: 2.60
IT: 2.73
FR: 2.84
ES: 3.26
DE: 2.78
AR: 2.84

# Case Study: SciELO Integration

- MT provides document based translations to SciELO thru the Bilingual Viewer (BV) interface
  - SciELO passes the document Url and the language selection

```
http://www.windowslivetranslator.com/BV.aspx?ref=AddIn&lp=pt_en&a=http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0085-56262008000200001&lng=pt&nrm=iso&tlng=pt&skpa=on
```

- BV takes care of the rest
  - Server interface
  - Progressive rendering

# Bilingual Viewer Interface

Windows Live™ Home Hotmail Spaces OneCare | MSN Search

Translate URL [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=s1519-69842008000200002&lng=en&nrn=iso&tng=es](http://www.scielo.br/scielo.php?script=sci_arttext&pid=s1519-69842008000200002&lng=en&nrn=iso&tng=es) English - Spanish

Original: English Go to original page Translated: Spanish Is this translation helpful?

articles articles search  
toc previous next author subject form home alpha

**Brazilian Journal of Biology**  
Print ISSN 1519-6984

**Braz. J. Biol. vol.68 no.2 São Carlos May 2008**  
doi: 10.1590/S1519-69842008000200002

**ECOLOGY**

**Spiders associated with *Psychotria carthagenensis* Jacquin. (Rubiaceae): vegetative branches versus inflorescences, and the influence of *Crematogaster* sp. (Hymenoptera, Formicidae), in South-Pantanal, Brazil**

**Aranhas associadas à *Psychotria carthagenensis* Jacq. (Rubiaceae): ramos vegetativos versus inflorescências, e a influência de *Crematogaster* sp. (Hymenoptera, Formicidae), no Pantanal Sul, Brasil**

**Services**

- Custom ser.
- Article in PDF
- Article in XML
- Article refere
- How to cite th
- Requests
- Cited by Scie
- Similar in S
- Automatic tra
- Send this art

**Diario brasileño de Biología**  
Imprimir ISSN 1519-6984

**Braz. j. Biol. vol.68 no.2 San Carlos de mayo de 2008**  
DOI: 10.1590/S1519-69842008000200002

**ECOLOGÍA**

**Arañas asociados con *Psychotria carthagenensis* Jacquin. (Rubiaceae): ramas vegetativa versus le y la influencia de *Crematogaster* SP Hymenoptera (Formicidae), en sur - Pantanal, Brasil**

**Aranhas associadas à *Psychotria carthagenensis* Jacq. (Rubiaceae): ramos vegetativos versus inflorescências, e a influência de *Crematogaster* SP (Hymenoptera, Formicidae), no Sul Pantanal, Brasil**

**Servicios**

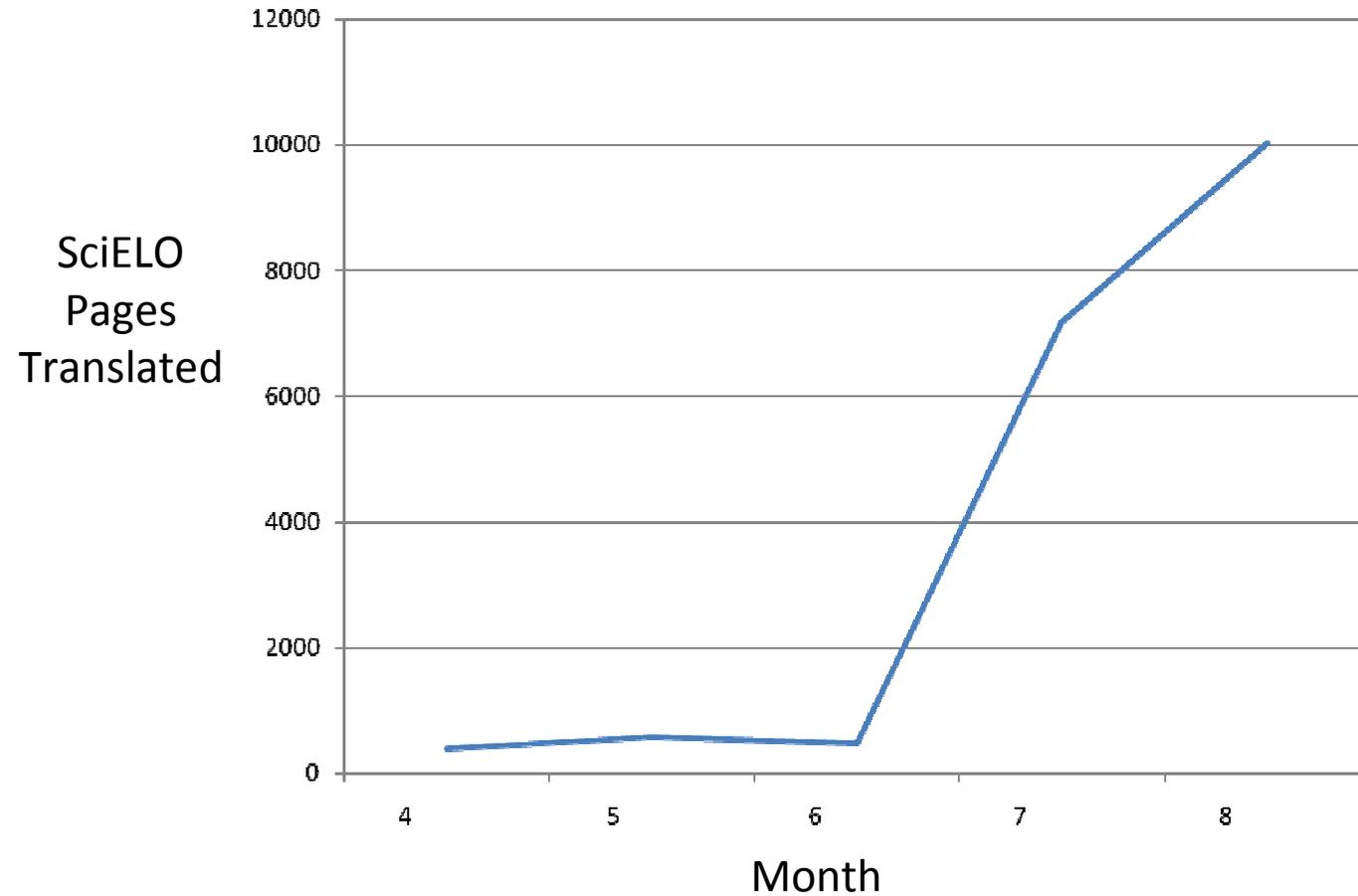
- Servicios pe
- Artículo en fc
- Artículo en fc
- Referencias
- Cómo citar a
- Solicitudes
- Citado por S
- Similares en
- Traducción e
- Enviar este a electrónico

© 2008 Microsoft | Privacy | Legal

Translation powered by Microsoft Translator

Translator Help | Help Central | Account | Feedback

# Adoption: SciELO Pages *Translated*



# New proposal: Query Translation

- Extend the reach of document search by translating the keywords before invoking the search itself
- Great fit for SciELO: Abstracts are already indexed in both the original article language and in English
- Suggested implementations:
  - Server side: SOAP call after tokenizing the query
  - Client side:
    - SOAP thru XmlHttpRequest
    - JSON

# Scenario: Spanish speaker user querying SciELO Brazil



revistas | artículos

alfab materia búsqueda autor materia búsqueda

## Colección de la biblioteca

su selección | enviar resultado | nueva búsqueda | config | fin de la página

Base de datos : article  
Búsqueda : caterpillars  
Referencias encontradas : 110 [refinar]  
Mostrando: 1 .. 10 en el formato [ISO 690]

página 1 de 11 va a la página 1 2 3 4 5 6 7 8 9 10 ▶ ▶▶

---

**1 / 110**

selecciona  
 para imprimir

Grego, Célia Regina, Vieira, Sidney Rosa and Lourenção, André Luiz **Spatial distribution of *Pseudaletia sequax* Franclmont in triticale under no-till management.** *Sci. agric. (Piracicaba, Braz.)*, Aug 2006, vol.63, no.4, p.321-327. ISSN 0103-9016

- [resumen en inglés](#) | [portugués](#) • [texto en inglés](#)

---

**2 / 110**

selecciona  
 para imprimir

Baptista, G.C. de, Parra, J.R.P. and Haddad, M. de L. **Toxicidade comparativa de lambda- cyalothrin à lagarta-da-soja, *Anticarsia gemmatalis* Hueb., 1818 (Lepidoptera, Noctuidae) e ao percevejo verde, *Nezara viridula* (L., 1758) (Hemiptera, Pentatomidae).** *Sci. agric. (Piracicaba, Braz.)*, Abr 1995, vol.52, no.1, p.183-188. ISSN 0103-9016

- [resumen en portugués](#) | [inglés](#) • [texto en portugués](#)

# Scenario: Spanish speaker user querying SciELO Brazil (new proposal)

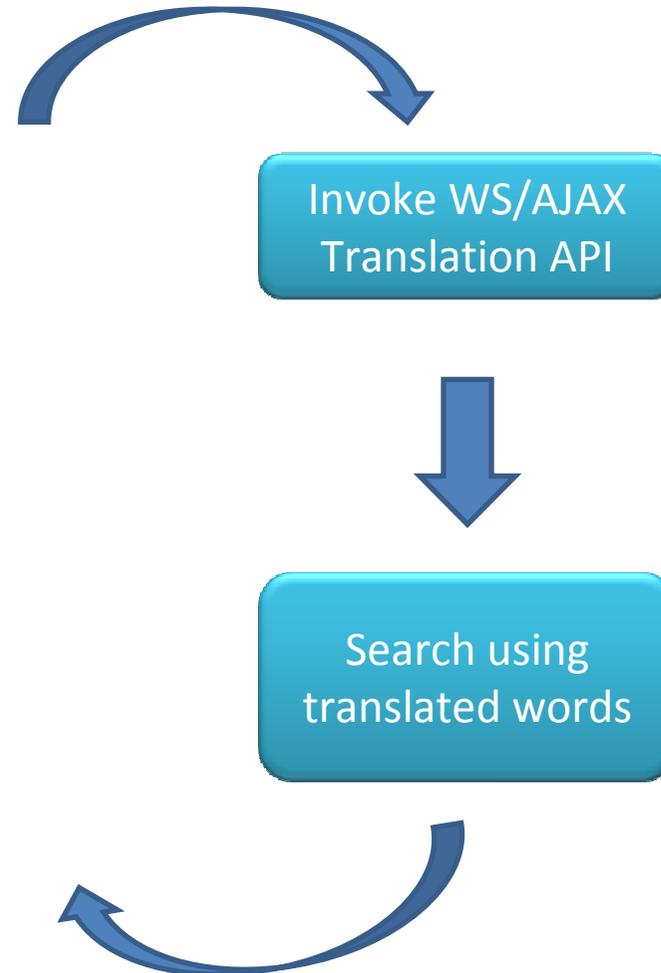
The image shows two screenshots of the SciELO Brazil search interface. The top screenshot shows the search input stage where the user has entered 'orugas' and selected the search options. The bottom screenshot shows the search results page with two articles listed.

**Top Screenshot (Search Input):**

- Logo: SciELO Brasil
- Navigation: revistas | artículos
- Search filters: alfab | materia | búsqueda | autor | materia | búsqueda
- Section: Colección de la biblioteca
- Base de datos: article
- Formulario libre: Formularios básicos
- Input: Ingrese una o más palabras: orugas
- Options:  Todas las palabras (AND)  Cualquier palabra (OR)
- Checkbox:  Traducir las palabras para inglés
- Buttons: config | búsqueda

**Bottom Screenshot (Search Results):**

- Logo: SciELO Brasil
- Navigation: revistas | artículos
- Search filters: alfab | materia | búsqueda | autor | materia | búsqueda
- Section: Colección de la biblioteca
- Buttons: su selección | enviar resultado | nueva búsqueda | config | fin de la página
- Base de datos: article
- Búsqueda: caterpillars
- Referencias encontradas: 108 [refinar]
- Mostrando: 1..10 en el formato [ISO 690]
- Page: página 1 de 11
- Navigation: va a la página 1 2 3 4 5 6 7 8 9 10 ▶ ▶▶
- Item 1 / 108:
  - selecciona
  - para imprimir
  - Grego, Célia Regina, Vieira, Sidney Rosa and Lourenção, André Luiz **Spatial distribution of *Pseudaletia sequax* Franckemont in triticale under no-till management.** *Sci. agric. (Piracicaba, Braz.)*, Aug 2006, vol.63, no.4, p.321-327. ISSN 0103-9016
  - resumen en inglés | portugués | texto en inglés
- Item 2 / 108:
  - selecciona
  - para imprimir
  - Baptista, G.C. de, Parra, J.R.P. and Haddad, M. de L. **Toxicidade comparativa de lambda- cyalothrin à lagarta-da-soja, *Anticarsia gemmatalis* Hueb., 1818 (Lepidoptera, Noctuidae) e ao percevejo verde, *Nezara viridula* (L., 1758) (Hemiptera, Pentatomidae).** *Sci. agric. (Piracicaba, Braz.)*, Abr 1995, vol.52, no.1, p.183-188. ISSN 0103-9016
  - resumen en portugués | inglés | texto en portugués



# New proposal: Query Translation

- Simple, well proven, scalable and stable interface to translation services
- Significant return on investment: the query results may include matches for *both* languages so the user get all matches in on dataset.

# So, what's next for MSR-MT?

- Support for more languages
- Support for more document types
- Interface with more Microsoft properties
- Improve translation quality
- Constantly working on performance improvements

# Thank you!

- [hribiero@microsoft.com](mailto:hribiero@microsoft.com)
- <http://translator.live.com>
- [tbot@hotmail.com](mailto:tbot@hotmail.com)
- <http://blogs.msdn.com/translation/>

Microsoft®  
**Translator**