

Have We Found the Holy Grail?

MT Summit 2003

Steve Richardson
Microsoft Research

NO

But we are at a pivotal juncture in the history of MT

- The viability of data-driven MT (DDMT) has been firmly established (ref AMTA 2002 theme)
 - DD (Statistical) MT systems are winning NIST competition (e.g., Och)
 - DDMT-based products have appeared (e.g., Language Weaver)
 - DDMT seems well-suited for a number of commercial domains & applications, more so with increasing globalization of various industries
 - Large-scale deployment of DDMT systems (e.g., MSR-MT)

Microsoft's Product Support Knowledge Base translated by MSR-MT, a DDMT system

<http://support.microsoft.com>

The screenshot shows a Microsoft Internet Explorer browser window displaying the Microsoft Support website in Spanish. The page title is "Microsoft Help and Support - Microsoft Internet Explorer". The address bar shows a search query for "estilo de viñeta en Powerpoint". The website header includes "Microsoft Soporte Online España" and navigation links like "Inicio", "Buscar", "Productos", "Descargas", and "Grupos de noticias". The main content area is titled "Buscar (Knowledge Base)" and shows search results for "estilo de viñeta en Powerpoint". The search results are displayed in a list format, with the first result being "PPT2000 Texto y Estilos de Viñeta Se Pierden después de Usted Copia y Cuadro de Texto de Pegado (271188)".

Microsoft Help and Support - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address Powerpoint&Query=estilo+de+vi%C3%B1eta+en+Powerpoint&Catalog=LCID%3D1034%26CDID%3DE5-KB%26PRODLISTSRC%3DON&QuerySource=gsfxSearch_Query Go Links

Microsoft Soporte Online España

Productos | Soporte | Buscar | Contáctenos

Inicio | Buscar | Productos | Descargas | Grupos de noticias

Enviar | Imprimir | Ayuda

Buscar (Knowledge Base)

estilo de viñeta en Powerpoint

Knowledge Base en español

Ver opciones

Quiero...

- Autoayuda
- Descargas y actualizaciones
- Soporte por producto
- Grupos de noticias
- Servicios de soporte
- Contactar con Microsoft

Soporte internacional

Optimize su búsqueda

Vista previa

Mostrar sólo resultados

Seleccione un producto Microsoft

- Todos los productos -

Buscar en los resultados

estilo de viñeta en Powerpoint

Ir

Resumen

Buscar en los resultados: 25 artículos

- [PPT2000 Texto y Estilos de Viñeta Se Pierden después de Usted Copia y Cuadro de Texto de Pegado \(271188\)](#) - Después de que copia y después de que pega un cuadro de texto que contiene una lista con viñetas, los estilos de texto y viñeta no se cambian a la hora de que promueve o a la hora de que degrada elementos de lista.
- [PPT Cómo Determina y/o Cambia el Estilo de Viñeta \(132189\)](#) - La fuente del primer carácter de una línea con viñeta determina la fuente utilizada para la viñeta cuando utiliza viñetas de Texto de Modo normal. Este comportamiento es normal para la función de viñeta en PowerPoint. Si hace clic en Viñeta en el menú Fo
- [WD98 Viñetas en Estilo Personalizado de Viñeta Cambian Numeradas A \(177208\)](#) - Cuando guarda aquél a un documento contiene un estilo personalizado con una viñeta como Word 6.0 o 95 .doc cambia el estilo numerado cuando vuelve a abrir el documento a. Por ejemplo NOTA este problema ocurre al guardar utilizar el convertidor Binario de
- [WD2000 Cambio de Estilo de Viñeta No Cambia Todas las Viñetas en Una de Jerarquía de Nivel \(204214\)](#) - Cuando utiliza viñetas en un documento y a continuación, cambia el estilo de una viñeta en una de la jerarquía bullet-style de nivel, las otras viñetas en este nivel no se actualizan automáticamente con el

Internet

start

5 W. C:\... Unti... 2 M... TR... 4 I... Micr... Micr... test... 10:28 PM

Role of Rule-based components

- Current example:
 - MSR-MT combines rule-based parsing with statistical terminology identification, example-based transfer, and machine-learned generation
- In the future:
 - Leverage tremendous amount of work and linguistic knowledge, using to generate training data, e.g.:
 - MSR-MT's Amalgam generation component, which learns to generate from corpus and Logical Forms produced by rule-based parser
 - Marker-based EBMT (Gough and Way), trained on output of Logomedia system
 - Some rule-based components are comprehensive and almost as easy to build (e.g., inflectional morphology) as their statistical counterparts

If we haven't yet found the Holy Grail, then what and where is it?

- One vision:
 - Ubiquitous MT: innumerable DDMT engines, spread across the Internet, each customized to produce high quality output relevant to a particular web site, application, or domain in a multitude of languages, each continuously adapting and learning from available sources (both human and machine)
 - Sponsored by a variety of owners/sources, based on many variations of (DDMT) technologies
 - Seamlessly accessible through automatic classification of text to be translated

How do we get there?

- Cost to create and maintain a customized MT system must drop!
 - Commodity economics
- This is the eventual promise of data-driven approaches
- Long time MT vendors like SYSTRAN are also focused on facilitating customization (see papers at this conference)
- Statistical methods are particularly well-suited to this task
 - Inherently adaptive
 - Adaptable – can be used to model anything, readily combinable with other methods, no limits on application
 - Robust – smoothing, etc.
- But any effective, adaptive methods are welcome

What's next?

- Lots of work: some interesting research as well as some tedious development
- Performance: in spite of ever-increasing computer speeds and capacities, we need faster, more efficient algorithms
- Accuracy: broader, more discriminating context sensitivity, modeling of additional information (linguistic & otherwise), seamless domain shifting
- Standardization of text formats, yielding more tractable training data
- Turn-key training that results in good quality
- Evaluation with increased diagnostic power
- System integration and scaling
- Human usability considerations