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)
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