- KM Concept
- KM Situation
- KM Obstacles
- KM Solutions
- LT Situation
- LT Obstacles
- LT Solutions
- Bitext.com

LangTech 2003
Paris FR

A new KM Approach based on the Neolithic Experience:

from foraging to farming Knowledge

(plus some notes on NLP and KM)

Antonio S. Valderrábanos
Bitext.com - The Bit and Text Company SL
http://www.bitext.com
asv@bitext.com

LangTech 2003 – Paris, France

Bitext.com

- **KM Concept**
- KM Situation
- KM Obstacles
- KM Solutions
- LT Situation
- LT Obstacles
- LT Solutions
- Bitext.com

What we understand by KM

- The context: Software and Productivity
  - –70-80 - **Personal productivity** applications (editors, spreadsheets, etc.)
  - –90-00 - **Group productivity** applications: current target for enterprises (and software industry)
  - **ERP, CRM, CMS/ECM, SCM, KM...**
- For the purpose of this presentation
  - –KM as a set of apps that targets group productivity
  - –with a focus on **highly-relevant** (**written**) **knowledge**
  - –and on two relevant phases: generation and consumption

- KM Concept
- **KM Situation**
- KM Obstacles
- KM Solutions
- LT Situation
- LT Obstacles
- LT Solutions
- Bitext.com

LangTech 2003
Paris FR

What's the situation now for KM apps

- There are high expectations
  - –For example: If I need to make a decision, KM will allow me to gather all relevant company knowledge, in a convenient format and time (regardless of its creator, language, form) so I can make the right decision (*SchlumbergerSema, SmartPractice*)
- and a formidable technical challenge
  - –Develop Knowledge Technologies and integrate them with Enterprise Management processes and practices
- Is it a feasible goal? What are the main obstacles?
- Similar situation to MT or NLP in the past?

- KM Concept
- KM Situation
- **KM Obstacles**
- KM Solutions
- LT Situation
- LT Obstacles
- LT Solutions
- Bitext.com

KM - Obstacles

- Focus at the end of the knowledge lifecycle
  - –KM concentrates on knowledge **consumption or use**: searching, categorization, etc.
- Contradiction between
  - –goal: KM **targets group productivity**
  - –means: KM **builds on personal productivity apps** that produce knowledge for personal inner-circle use (editors)
- •Knowledge foragers (or hunters-gatherers)
  - –Paradigmatic example: **search engine**
  - –Knowledge grows elsewhere and it's hunted for, individually

KM - Solutions (1 of 2)

- Focus at the beginning of the knowledge lifecycle
  - Change in focus: from consumption to **creation**
- Creation of critical knowledge should be done
  - in a **single multiuser application** (CMS)
  - according to **group rules** (not as an individual activity)
  - probably keeping a link between form and meaning
- Knowledge farming
  - knowledge is grown under control
  - currently complex tasks (like searching) become trivial

- KM Concept
- KM Situation
- KM Obstacles
- **KM Solutions**
- LT Situation
- LT Obstacles
- LT Solutions
- Bitext.com

KM – Solutions (2 of 2)

- Knowledge farming: early hints and adopters
  - **Simple techniques are already in use**
  - Templates and forms: central management of *n* authors
  - Controlled terminologies: Tech Authoring, TM tools
  - Doc structure: Acrobat Bookmark, MSWord Doc Maps

    Martin Langham, Bloor Research: 50% have structure

  - **More sophisticated techniques are taking off**
  - Controlled language (LTI-CMU - US, Caterpillar)
  - Conceptual Authoring (ITRI-UBrighton - UK, PILLS)
  - High formalization levels of written knowledge
- An IG is in place (SLBS, DFKI, EADS, XRCE...)

- KM Concept
- KM Situation
- KM Obstacles
- KM Solutions
- **LT Situation**
- LT Obstacles
- LT Solutions
- Bitext.com

What's the situation now for LT tech

- Widely-used tools don't use (even basic) LT!

- The Google case
  - –handling of "weird" characters (á, ü, ç)
    - inconsistent documentation (English vs. Spanish Help)
    - changing attitude
  - –spelling algorithm for user queries
    - based on string frequency, no language knowledge
    - error-prone: **correct words are reported as "incorrect"**
      - –EN query: ***nuked*** - Did you mean ***naked***
      - –ES query*:* ***desnucar*** - Did you mean ***desnuda***

- But it may change: Applied Semantics (AdSense)

- KM Concept
- KM Situation
- KM Obstacles
- KM Solutions
- LT Situation
- **LT Obstacles**
- LT Solutions
- Bitext.com

LT - Obstacles

- Basic resources are scarce-costly-expensive
  - Slow development cycles
  - Complex pricing and licensing schemes (early ROI)
  - *Success of statistical approaches (Autonomy)*
- Atomic approach to market penetration
  - Growing (but short) number of **small players**
  - Aiming at developing **full (and similar) solutions**
    - intelligent search and indexing
- Strong focus on new developments (NLU)
  - rather than on deployment of mature developments
  - doc categorization vs. query expansion (verbs)

- KM Concept
- KM Situation
- KM Obstacles
- KM Solutions
- LT Situation
- LT Obstacles
- **LT Solutions**
- Bitext.com

LangTech 2003
Paris FR

NLP - Solutions (1 of 2)

- Exploit the KM boost
  - –KM community: productivity-driven, not research-driven; and **well integrated in enterpise structure**
  - –NLP community: the opposite!
- Try a different market penetration strategy
  - –**Externally**
  - –tighten integration with KM players (Plumtree, OpenText)
  - –**Internally**
  - –develop cooperation agreements (diff languages)
  - –merge partial solutions (search and classification)

  - –**FP6 is forcing this move!**

LT - Solutions (2 of 2)

- Use a different strategy for resource development
  - –public funding at 100%, not 50% (tenders)
  - –make them publicly available for research (WordNet)
  - –develop reasonable licensing and pricing schemes
  - –provide framework for copyright protection
- This could be a good moment
  - –70% budget increase for FP7 (KTWeb)
  - –Requested by European Parliament
  - –due to EU enlargement

- KM Concept
- KM Situation
- KM Obstacles
- KM Solutions
- LT Situation
- LT Obstacles
- LT Solutions
- **Bitext.com**

LT - What we are doing at Bitext.com

- Develop a strategy focused on integration
  - –building basic NLP services
    - spelling, query expansion, NLI (shallow analysis)...
  - –with a modular and cost-effective approach
  - –developed for market players (not for end-users)
    - SchlumbergerSema sae, dtSearch Inc., Atril SL, FutureSpace SA (RENFE), iSOCO SA, Carrot SL...
  - –cooperating with research institutions
    - UPM, USev, UPF, RALI (UdeM)
  - –combining R&D and D
    - applied research: LIQUID, TT2, ALLES (IST funding)

Thank you for your attention

Antonio S. Valderrábanos
Bitext.com - The Bit and Text Company SL
http://www.bitext.com
asv@bitext.com

Bitext.com