

MT Summit VIII - generelle tendenser

Bente Maegaard

Seminar om de væsentligste nyheder fra MT Summit

15. november 2001

MT Summit VIII



Hvor: Santiago de Compostela, Galicia, 18-22 sep.
2001

MT Summit deltagere

- **Brugere**
- **Forskere**
- **Udviklere**
- **Producenter**
- **Konsulenter, andre med en (professionel) interesse i MT**

Formålet med MT Summit



- **Høre om den nyeste udvikling**
 - på markedet
 - i forskningen
- **Diskutere på et professionelt niveau**
- **Skabe synergi**
- **Skabe et mødested, nye partnerskaber**
- **Med særligt fokus på brugerne**

MT Summit - hvad er MT?



Alle typer af hjælpemidler til oversættelse

- **Ikke blot rendyrket MT**
- **Også oversættelseshukommelse**
- **Termbaser**
- **Workflowsystemer**
- **Integrerede systemer**
- **Ordbøger, terminologi**

Programmet: workshopper



4 workshopper

- **Workshopper**
 - **EBMT**
 - **MT2010 - Roadmap for MT**
 - **MT Evaluation**
 - **Teaching MT**
- **Åbent møde om terminologi (aflyst)**

Oversigt over typen af indlæg

- **MT-systemer 11**
- **MT i praksis 6**
- **Oversættelse integreret (i workflow, på web) 4**
- **Udvikling af ordbøger til MT 3**
- **Tosprogede ordbøger til MT 3**
- **Oversættelseshukommelse, alignment 3**
- **MT og sproglige ressourcer (inkl. standarder)**

Konferenceprogrammet



Oversigt over typen af indlæg (fortsat)

- **MT-metoder 11**
- **Generering 7**
- **Morfologi til MT 5**
- **Evaluering 11**

- **4 paneldiskussioner**
 - **Brugere; Evaluering; MT i de 3 regioner; Fremtiden**

Brugerpanel



Panel: User needs

Brugere

Annelise Bech

Lingtech

Jennifer Brundage

SAP

Laurie Gerber

Language Technology
Broker

Producenter

Jean Senellart

Systran

Francis Bond

NTT

Marjorie Leon

PAHO

Spørgsmålene til brugerne



Brugerne fremlagde deres svar på:

- **Hvad er de vigtigste FORDELE ved at bruge MT?**
- **Hvad er de vigtigste PROBLEMER?**
- **Og hvad ønsker du dig derfor af fremtiden?**

Fordele



De vigtigste fordele

- **Kvalitet og konsistens i terminologi**
- **Økonomi**
- **Højere produktivitet, sparer tid**
- **Genbrug af materiale**
- **Videnopbygning i firmaet, støtte til nyansatte**
- **Altid parat**
- **Kan håndtere store mængder**
- **Fortrolighed - ved opgaver for kunder**

De vigtigste problemer

- **Systemer fra forskellige leverandører er ikke kompatible, ingen åben API**
- **For hyppige nye versioner**
- **Alignment-værktøjer er ikke gode nok**
- **Man deles ikke om store ordbøger**
- **Ingen generelle udvekslingsformater**
- **Dårlige værktøjer til ordbogsproduktion**
- **Savner samarbejde mellem leverandørerne**
- **Ingen stærk brugerbevægelse**

Problemer - resultater af diskussionen eller indlæg på selve konferencen



De kursiverede er nogen indstillet på at gøre noget ved

- ***Systemer fra forskellige leverandører er ikke kompatible, ingen åben API***
- For hyppige nye versioner
- ***Alignment-værktøjer er ikke gode nok***
- Man deles ikke om store ordbøger
- ***Ingen generelle udvekslingsformater***
- ***Dårlige værktøjer til ordbogsproduktion***
- Savner samarbejde mellem leverandørerne
- ***Ingen stærk brugerbevægelse***

Gøre lukkede systemer mere åbne



Systrans 'New Generation Translation Systems'

I forbindelse med udviklingen af en række nye sprogpar har Systran arbejdet med systemdesign.

Engelsk-ungarsk, engelsk-polsk, engelsk-arabisk, ungarsk-fransk, polsk-fransk

Der vil blive behov for et kolossalt antal sprogpar

Gøre lukkede systemer mere åbne



Systrans 'New Generation Translation Systems'

- **Behov for**
 - **hurtig udvikling af nye sprog og sprogpar**
 - **individ- og kundetilpassede ordbøger, grammatikker mv.**
 - **ikke bare MT, men fx flersproget indeksering, forfatterværktøjer, integration med oversættelseshukommelse**
 - **nemt at udveksle terminologidata, integrere værktøjer til produktion af bl.a. ordbøger**

Gøre lukkede systemer mere åbne



Systrans 'New Generation Translation Systems'

Metoder

Modularitet

- **Analyse-transfer-syntese**

Standardudvekslingsformat (OLIF)

Eksterne komponenter

- **fx 'ekstern analyse'-transfer-syntese**
- **Intuitivt kodningsværktøj til ordbøger**

Alignment ikke god nok



Flere foredrag om nye metoder i alignment

- **Man har behov for gode alignmentværktøjer når man arbejder med ordbøger**
- **Her er det ord eller udtryk, ikke sætninger, der skal 'parres'**
- **Et godt alignmentværktøj kan give vældig gode forslag til indgange i en tosproget ordbog**
- **Forslag om at bruge 'cognates' (ord der ligner hinanden, Parliament-parlament) som faste holdepunkter**

OLIF, SALT

- **OLIF**

- OLIF, the Open Lexicon Interchange Format, is a user-friendly vehicle for exchanging terminological and lexical data. OLIF overholder XML-standarden.

- **Deltagere**

- SAP, Europa-Kommissionen, Microsoft, Lionbridge, Lotus, Xerox, Trados, Saillabs, Systran, Logos, IBM, DFKI, IAI

- **Adresse: www.olif.net**

OLIF, SALT

- **SALT**

- Projekt (EU-USA). Kombinerer OLIF (for leksika) og MARTIF (for terminologi). Formålet er at skabe et fælles databaseformat, XLT, der integrerer ordbogs- og termdata.

- **Deltagere**

- Univ. i Wien, Köln, Bolzano, Surrey, Brigham Young, Kent State, LORIA, IAI

Adresse: www.ttt.org/SALT/

Ordbøger til MT

- **Systran - jf. tidligere**
- **Logos - en gennemgang af processen**
 - **Bygger meget på eksisterende materiale**
 - **OLIF**
- **Univ. Brighton**
 - **WASP-bench, - workbench til MT-ordbogsproduktion, entydiggørelse, kontekster**

Brugerbevægelse



Kan man skabe et IAMT netværk af brugere?

- **En stærk interesse udvikledes**
 - **Laurie Gerber**
 - **Jennifer Brundage**
- **Der er klart fælles interesser**
- **Man kunne få en side i MT News International**
- **Man kunne oprette en hjemmeside (IAMT)**
- **Kan man holde det i gang?**

Brugernes ønsker



Brugerpanelet blev også bedt om deres ønsker for fremtiden

- **De fleste ønsker er dækket af problemformuleringerne**
- **Udveksling af terminologi**
- **Automatisk sprogidentifikation**
- **Kontrolleret sprog, integreret med MT**

Konklusioner



Der er gang i maskinoversættelse

- **Forskning i metoder**
- **Udvikling af hjælpeværktøjer - ordbog, alignment**
- **Brug af store tekstmængder**
- **Flere sprog og sprogpar**