ubiquity:
designing a multilingual natural language interface

mitcho
Michael Yoshitaka ERLEWINE

2009 SIGIR Workshop on Information Access in a Multilingual World
Boston, July 23rd, 2009
ubiquity: designing a multilingual natural language interface

• focus on multilingual delivery and interfaces
• reconsidering the text-based interface
• localizing ubiquity
ubiquity: designing a multilingual natural language interface

- focus on multilingual delivery and interfaces
- reconsidering the text-based interface
- localizing ubiquity
INTERFACE

USER

INFORMATION

English
Deutsch
Español
Français
Português
Русский
Italiano
Polski
Türkçe
日本語
中文
this talk
EQUIPOTENT MULTILINGUAL INTERFACES

- not just translating content or localizing interfaces
- ensuring equal access to information
- ensuring equal generativity and potential of the platform
EQUIPOTENT MULTILINGUAL INTERFACES

- not just translating content or localizing interfaces
- ensuring equal access to information
- ensuring equal generativity and potential of the platform
EQUIPOTENT MULTILINGUAL INTERFACES

- not just translating content or localizing interfaces
- ensuring equal access to information
- ensuring equal generativity and potential of the platform
“We create communities of people involved in making the Internet experience better for all of us... The Internet is a global public resource that must remain open and accessible.”

—Mozilla Manifesto
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afrikaans</td>
<td>Afrikaans</td>
</tr>
<tr>
<td>Shqip</td>
<td>Shqip</td>
</tr>
<tr>
<td>العربية</td>
<td>العربية</td>
</tr>
<tr>
<td>असमीया</td>
<td>অসমীয়া</td>
</tr>
<tr>
<td>Euskara</td>
<td>Euskara</td>
</tr>
<tr>
<td>Беларуская</td>
<td>Беларуская</td>
</tr>
<tr>
<td>भारत</td>
<td>भारत</td>
</tr>
<tr>
<td>Български</td>
<td>Български</td>
</tr>
<tr>
<td>català</td>
<td>català</td>
</tr>
<tr>
<td>тх</td>
<td>тх</td>
</tr>
<tr>
<td>ਖ਼ਪਤਤ</td>
<td>ਖ਼ਪਤਤ</td>
</tr>
<tr>
<td>Hrvatski</td>
<td>Hrvatski</td>
</tr>
<tr>
<td>Čeština</td>
<td>Čeština</td>
</tr>
<tr>
<td>Dansk</td>
<td>Dansk</td>
</tr>
<tr>
<td>Nederlands</td>
<td>Nederlands</td>
</tr>
<tr>
<td>English (British)</td>
<td>English (British)</td>
</tr>
<tr>
<td>English (US)</td>
<td>English (US)</td>
</tr>
<tr>
<td>Esperanto</td>
<td>Esperanto</td>
</tr>
<tr>
<td>Eesti keel</td>
<td>Eesti keel</td>
</tr>
<tr>
<td>suomi</td>
<td>suomi</td>
</tr>
<tr>
<td>Français</td>
<td>Français</td>
</tr>
<tr>
<td>Frysk</td>
<td>Frysk</td>
</tr>
<tr>
<td>Galego</td>
<td>Galego</td>
</tr>
<tr>
<td>ქართული</td>
<td>ქართული</td>
</tr>
<tr>
<td>Deutsch</td>
<td>Deutsch</td>
</tr>
<tr>
<td>Ελληνικά</td>
<td>Ελληνικά</td>
</tr>
<tr>
<td>गोवरणी</td>
<td>गोवरणी</td>
</tr>
<tr>
<td>Kannada</td>
<td>Kannada</td>
</tr>
<tr>
<td>한국어</td>
<td>한국어</td>
</tr>
<tr>
<td>Kurdî</td>
<td>Kurdî</td>
</tr>
<tr>
<td>Latviešu</td>
<td>Latviešu</td>
</tr>
<tr>
<td>lietuvių kalba</td>
<td>lietuvių kalba</td>
</tr>
<tr>
<td>Македонски</td>
<td>Македонски</td>
</tr>
<tr>
<td>मराठी</td>
<td>मराठी</td>
</tr>
<tr>
<td>Norsk bokmål</td>
<td>Norsk bokmål</td>
</tr>
<tr>
<td>Norsk nynorsk</td>
<td>Norsk nynorsk</td>
</tr>
<tr>
<td>occitan (lengadocian)</td>
<td>occitan (lengadocian)</td>
</tr>
<tr>
<td>فارسی</td>
<td>فارسی</td>
</tr>
<tr>
<td>Polski</td>
<td>Polski</td>
</tr>
<tr>
<td>Português (do Brasil)</td>
<td>Português (do Brasil)</td>
</tr>
<tr>
<td>Português (Europeu)</td>
<td>Português (Europeu)</td>
</tr>
<tr>
<td>ਪੰਜਾਬੀ</td>
<td>ਪੰਜਾਬੀ</td>
</tr>
<tr>
<td>română</td>
<td>română</td>
</tr>
<tr>
<td>Русский</td>
<td>Русский</td>
</tr>
<tr>
<td>српски</td>
<td>српски</td>
</tr>
<tr>
<td>සිංහල</td>
<td>සිංහල</td>
</tr>
<tr>
<td>Slovenčina</td>
<td>slovenčina</td>
</tr>
<tr>
<td>slovensko</td>
<td>slovensko</td>
</tr>
<tr>
<td>Español (de Argentina)</td>
<td>Español (de Argentina)</td>
</tr>
<tr>
<td>Español (de España)</td>
<td>Español (de España)</td>
</tr>
<tr>
<td>Svenska</td>
<td>Svenska</td>
</tr>
<tr>
<td>Telugu</td>
<td>Telugu</td>
</tr>
<tr>
<td>ไทย</td>
<td>ไทย</td>
</tr>
<tr>
<td>Українська</td>
<td>Українська</td>
</tr>
<tr>
<td>Tiếng Việt</td>
<td>Tiếng Việt</td>
</tr>
<tr>
<td>Cymraeg</td>
<td>Cymraeg</td>
</tr>
</tbody>
</table>
ubiquity: designing a multilingual natural language interface

• focus on multilingual delivery and interfaces
• reconsidering the text-based interface
• localizing ubiquity
BUILDING A SCALABLE INTERFACE PARADIGM
BUILDING A SCALABLE INTERFACE PARADIGM

The web is offering more and more functionality and API’s...
BUILDING A SCALABLE INTERFACE PARADIGM

The web is offering more and more functionality and API’s...

...but how do you use them?
BUILDING A SCALABLE INTERFACE PARADIGM

The web is offering more and more functionality and API’s...

...but how do you use them?

• Following the GUI paradigm...
BUILDING A SCALABLE INTERFACE PARADIGM

The web is offering more and more functionality and API’s…

…but how do you use them?

• Following the GUI paradigm…
  • add more buttons?
BUILDING A SCALABLE INTERFACE PARADIGM

The web is offering more and more functionality and API’s...

...but how do you use them?

• Following the GUI paradigm...
  • add more buttons?
  • add more menus?
ACM interactions 2008

The Linguistic Command Line

Aza Raskin
Humanized | aza@humanized.com
BASH
Aza Raskin on text-based interfaces

“...maybe this confusion isn’t the fault of command-line interfaces in general; maybe it’s just the command lines we’re used to. If commands were memorable, and their syntax forgiving, perhaps we wouldn’t be so scared to reconsider these interface paradigms.”
Aza Raskin on text-based interfaces

“...maybe this confusion isn’t the fault of command-line interfaces in general; maybe it’s just the command lines we’re used to. If commands were memorable, and their syntax forgiving, perhaps we wouldn’t be so scared to reconsider these interface paradigms.”
a natural language interface
[DEMO 1]

inserting a map into email
(video not included in PDF)
without ubiquity
without ubiquity
without ubiquity
without ubiquity
[DEMO 1] inserting a map into email: one step with *ubiquity* (video not included in PDF)
[DEMO 2]

translating and emailing web content
(video not included in PDF)
東京ビッグサイト(東京都江東区)で開催中の「LinuxWorld Expo/Tokyo 2008」(28日～30日)において、「オープンソース焼きそば」なるものが配布されている。野村総合研究所(NRI)が開発したもので、同社ブースで入手可能。ヤケドしそうな熱湯をかけて3分後、湯切りを実施してからソースをからめることで完成。Linuxなど知らないうーザーでも美味しく食せるという。

今年で10周年を迎えるLinuxWorld Expo(IDGジャパン主催)。最新のオープンソース・テクノロジーに触れられる国内最大級のイベントだ。出展企業各社によるさまざまな製品やソリューションが紹介され、趣味を凝らしたノベルティグッズでアピールするブースも多く目につく。そん
without ubiquity
without ubiquity
without ubiquity
without ubiquity
translating and emailing web content: fast and straightforward with ubiquity
(video not included in PDF)
GOAL:

HOW?
GOAL: accomplish more on the web

HOW?
GOAL: accomplish more on the web

HOW? connect the web with language
VERBS*
ubiquity: designing a multilingual natural language interface

- focus on multilingual delivery and interfaces
- reconsidering the text-based interface
- localizing ubiquity
The goal: “natural syntax”

- Users must expect that not all natural inputs are understandable.

- But the parser must not add restrictions which go against the user’s natural syntax.
The goal: “natural syntax”

- Users must expect that not all natural inputs are understandable.

- But the parser must not add restrictions which go against the user’s natural syntax.
LOCALIZING UBIQUITY

**EX:** The following must behave equivalently:

(Japanese, a scrambling language)

- 太郎にボールを投げる
  Tarō-ni ball-o nager=o
  Tarō-DAT ball-ACC throw=IMPER
- ボールを太郎に投げる
**EX:** The following must behave equivalently:

(Japanese)

- ...投げろ nager=о (**IMPERATIVE**)
- ...投げる nager=u (**INFINITIVE**)
- ...投げて nage=te (**LINKING FORM**)
Putting theoretical linguistic research and insights into practice

- Writing individual language parsers is difficult and time-intensive
- A Principles and Parameters model
- One universal parser with minimal language settings
Putting theoretical linguistic research and insights into practice

- Writing individual language parsers is difficult and time-intensive

- A Principles and Parameters model

- One universal parser with minimal language settings
Putting theoretical linguistic research and insights into practice

- Writing individual language parsers is difficult and time-intensive

- A Principles and Parameters model

- One universal parser with minimal language settings
LOCALIZING UBIQUITY

- Putting theoretical linguistic research and insights into practice
  - Writing individual language parsers is difficult and time-intensive
  - A Principles and Parameters model
  - One universal parser with minimal language settings
parser

(animation not included in PDF)
花子にケーキを送って
花子にケーキを送って
花子に ケーキを 送って
parser

(Animation not included in PDF)

花子に ケーキを 送って

N  N  V

goal  object
花子に ケーキを 送って

N  N  V

goal  object

V( goal, object )
LOCALIZING UBIQUITY

Argument identification via semantic role

CmdUtils.CreateCommand({
    names: ["email", "mail"],
    arguments: [
        {role: "object", nountype: noun_arb_text},
        {role: "goal", nountype: noun_type_contact}
    ],
...
LOCALIZING UBIQUITY

Argument identification via semantic role

CmdUtils.CreateCommand({
    names: ["email", "mail"],
    arguments: [
        {role: "object", nountype: noun_arb_text},
        {role: "goal", nountype: noun_type_contact}
    ],
    ...
)}
LOCALIZING UBIQUITY

Argument identification via semantic role
LOCALIZING UBIQUITY

Argument identification via semantic role
**LOCALIZING UBIQUITY**

Argument identification via semantic role

<table>
<thead>
<tr>
<th>role</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
LOCALIZING UBIQUITY

Argument identification via semantic role

<table>
<thead>
<tr>
<th>role</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>ø</td>
<td>を o</td>
</tr>
</tbody>
</table>
## LOCALIZING UBIQUITY

Argument identification via semantic role

<table>
<thead>
<tr>
<th>role</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>Ø</td>
<td>を o</td>
</tr>
<tr>
<td>goal</td>
<td>to</td>
<td>に ni, へ e</td>
</tr>
</tbody>
</table>
# LOCALIZING UBIQUITY

Argument identification via semantic role

<table>
<thead>
<tr>
<th>role</th>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>$\emptyset$</td>
<td>を o</td>
</tr>
<tr>
<td>goal</td>
<td>to</td>
<td>に ni, へ e</td>
</tr>
<tr>
<td>location</td>
<td>at, on</td>
<td>で de</td>
</tr>
</tbody>
</table>
ubiquity in Japanese: “hello to Spanish” (video not included in PDF)
LOCALIZING UBIQUITY

- Support for various linguistic structures
  - Equal support for verb-initial and -final languages
  - Smart argument-first suggestions
LOCALIZING UBIQUITY

Minimal language descriptions

branching: 'left',
usespaces: false,
joindelimiter: '',
anaphora: ['これ', 'それ', 'あれ'],
roles: [
  {role: 'object', delimiter: 'を'},
  {role: 'goal', delimiter: 'に'},
  {role: 'source', delimiter: 'から'},
  {role: 'position', delimiter: 'で'},
  {role: 'alias', delimiter: 'として'},
]
LOCALIZING UBIQUITY

- An equipotent multilingual platform *par excellence*
- Equivalent underlying parse structures
- Commands via semantic role: write once, use in any language
- Easy to add new languages
LOCALIZING UBIQUITY

- An equipotent multilingual platform *par excellence*
  - Equivalent underlying parse structures
  - Commands via semantic role: write once, use in any language
  - Easy to add new languages
LOCALIZING UBIQUITY

- An equipotent multilingual platform *par excellence*
  - Equivalent underlying parse structures
  - Commands via semantic role: write once, use in any language
  - Easy to add new languages
ubiquity 0.5
with Parser 2

Language settings:
- Català
- Dansk
- English
- Español
- Français
- Italiano
- 日本語
- Português
- Svenska
- 中文
ubiquity 0.5
with Parser 2

Català
Dansk
English
Português
日本語
ubiquity: designing a multilingual natural language interface

- **Conclusion**
  - broad and equal access to information requires equipotent multilingual interfaces
  - **ubiquity**: an equipotent platform *par excellence*
  - challenging internationalization possible through application of linguistic research
ubiquity: designing a multilingual natural language interface

• **Conclusion**

  • broad and equal access to information requires equipotent multilingual interfaces

• **ubiquity**: an equipotent platform *par excellence*

• challenging internationalization possible through application of linguistic research
ubiquity: designing a multilingual natural language interface

• Conclusion
  • broad and equal access to information requires equipotent multilingual interfaces
  • ubiquity: an equipotent platform *par excellence*
  • challenging internationalization possible through application of linguistic research
Thank you!

Questions?