

The multilingual web

Felix Sasaki

DFKI / University of Appl. Sciences Potsdam

W3C German-Austrian Office

felix.sasaki@dfki.de

Slides:

<http://www.sasakiatcf.com/felix/publications/sasaki-webtechcon2010.pdf>

About me

- Studied Japanese and Linguistics in Germany and Japan
- PhD in Computational Linguistics with a focus on Web technology & multilingual data
- 2005-2009: Work in Japan within the W3C Internationalization Activity
- Since 2009: Professor at Univ. of Appl. Sciences Potsdam / Manager of the W3C German-Austrian Office
- Since Autumn 2010: Senior Researcher at DFKI (German Research Center for Artificial Intelligence)

About W3C Offices

- A contact point for those who
 - Don't know W3C very well (yet)
 - Want to ask specific questions like “Who is working on topic ABC ...”
 - Are considering new web-related standardization work and are wondering where the best place to do that might be
- So for any questions related to the topics mentioned above, please don't care to bother me too 😊

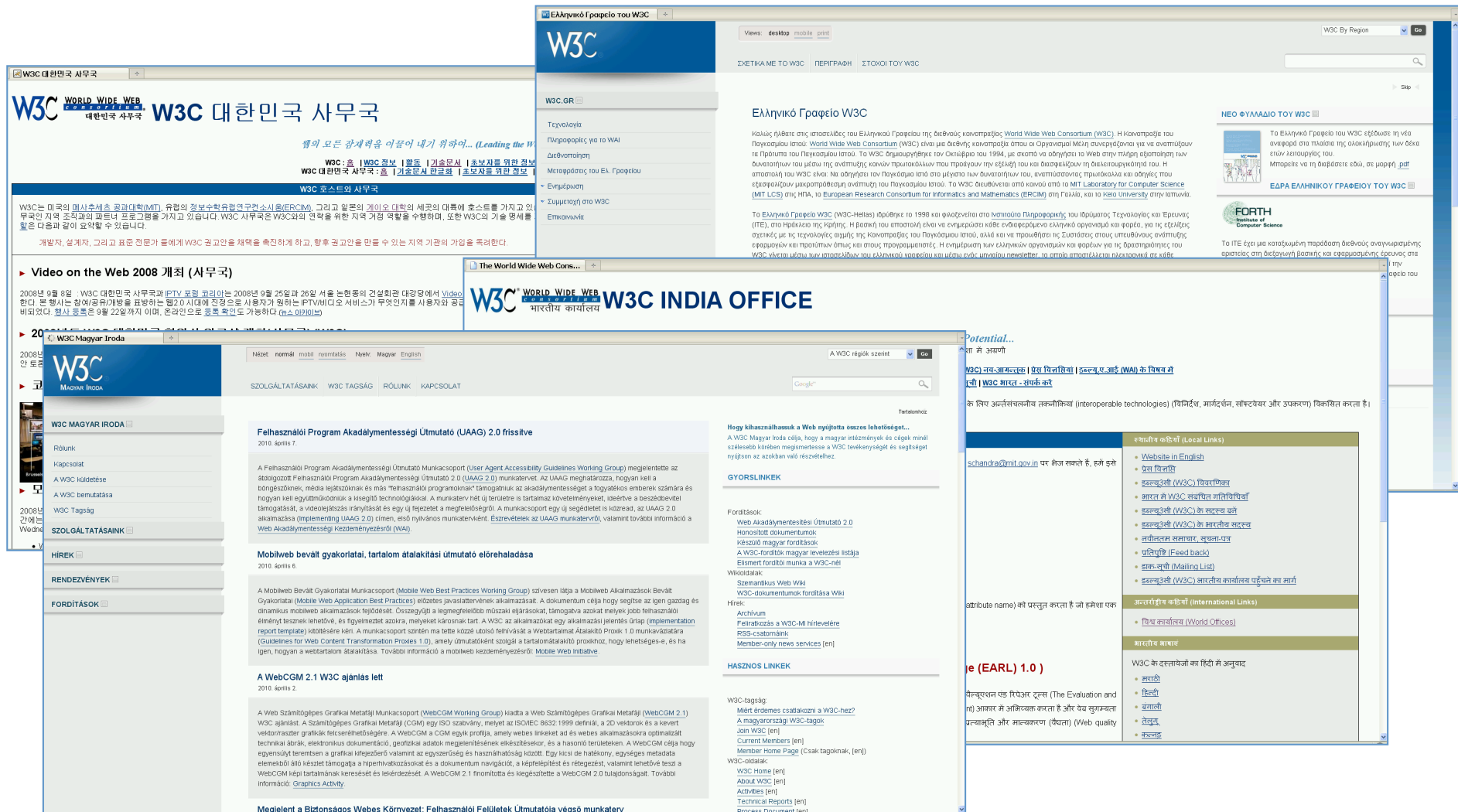
Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

Overview

- Multilingual Web – what's that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

The Multilingual Web: Content in many languages and scripts



Localized Services

The image displays four overlapping screenshots of the W3C CSS Validation Service interface, demonstrating localization into different languages:

- English (top):** Shows the header "W3C CSS Validation Service" with a list of languages including Deutsch, English, Español, Français, 한국어, Italiano, Nederlands, 日本語, Polski, Português, Русский, العربية, Svenska, Български, Українська, Čeština, Romanian, and 简体中文. The main text reads "CSS Validation Service" and "بررسی شویبه نامه های آنتشاری (CSS) و اسناد (HTML) دارای شویبه نامه".
- Korean (second):** Shows the header "W3C CSS Validation Service" with the same language list. The main text reads "CSS Validation Service" and "캐스캐이딩 스타일 시트 (CSS) 와 스타일시트를 포함한 (X)HTML 문법 검사". Below the header are buttons for "URI", "파일 업로드", and "직접 입력", and a "URI 검사" label.
- German (third):** Shows the header "W3C CSS Validation Service" with the same language list. The main text reads "CSS Validation Service" and "Cascading Style Sheets (CSS) und (X)HTML-Dokumente mit CSS validieren". Below the header are buttons for "per URI", "per File-Upload", and "per Direkteingabe".
- Spanish (bottom):** Shows the header "W3C CSS Validation Service" with the same language list. The main text reads "CSS Validation Service" and "Verifica Hojas de Estilo en Cascada (CSS) y documentos (X)HTML con hojas de estilo". Below the header are buttons for "mediante URI", "mediante Carga de Archivo", and "mediante Entrada directa". The main content area is titled "Validar mediante URI" and includes a text input field for "Dirección:" and a "Check" button. A note at the bottom states: "Nota: Si deseas validar tu hoja de estilo CSS incrustada en un documento (X)HTML, deberías antes [comprobar que el \(X\)HTML utilizado es válido](#)."

Devices everywhere, for everybody

Home

Main Menu

- Home
- News
- Events
- Workshops
- Press Room
- Public Documents
- Key People
- Search

Themes

- Mobile Web for Social Development
- Low cost information access devices
- Low cost broadband access and infrastructure
- Dissemination and outreach

Mobile Web for Social Development (MW4D)

The MW4D Interest Group explores how to use the potential of Information and Communication Technologies (ICTs) on Mobile phones as a solution to bridge the Digital Divide and provide minimal services (health, education, governance, business,...) to rural communities and under-privileged populations.

Home Skip Navigation Site Map Mobile View Search Subscribe Join the Mailing List Join the Discussion Donate Now

WORLD WIDE WEB FOUNDATION Vision Programs Projects Get Involved News About Contact

Advance the Web. Empower People.

We envision a Web that empowers all people, including billions excluded today, by providing new opportunities for creativity, collaboration, teaching, learning, enterprise and a better life. To reach this goal, the Web Foundation will put the power of the Web into the hands of people around the world through effective, high-impact programs.

The Foundation works to remove barriers to creating and consuming Web content, accelerate growth of the Web, understand how the Web works, advance Web technology, and ensure a free and open Web that empowers people to bring about social and economic change.

[More About Our Projects](#)
[Get Involved](#)

[Advance the Web. Empower People.](#)
[Regreening in Africa](#)
[Visits in Africa](#)
[One Web](#)
[Empowering Youth](#)
[More stories](#)

"Web for All" Conference Keynote

Apr 26, 2010 · Steve Bratt

The Web is not fully accessible to perhaps 2 billion, including people with:

- Disabilities
- Challenges due to aging
- Literate in poorly-supported languages
- Low literacy
- Little experience with technology

Today's Web Largely Excludes 1-2 Billion People

People with disabilities, aging, reading poorly-supported languages, with low literacy and/or lack of technology experience.

Today and tomorrow, in Raleigh, North Carolina, USA, the 7th International Cross-Disciplinary Conference on Web Accessibility is taking place. Most call this event Web for All or W4A. The focus this year is "Developing Regions: Common Goals, Common Problems", which is perfect fit for the Web Foundation's initial focus. I had the honor of delivering the opening keynote, entitled, "Breaking Barriers to a Read/Write Web that Empowers All" (PDF, PPT).

[Read More](#)

Donate

Our future programs depend on your support. Select your preferred donation amount to offer monetary contributions by credit card or PayPal and give the gift of empowerment.

Donate Now

Volunteer or Apply

Whether you're a scholar, student or practitioner, the Web Foundation can use your help to advance the Web and empower all people. Browse our current Volunteer and Employment Opportunities and lend a hand to humanity.

Current Opportunities

Join the Discussion

A free and open Web starts with you. Share your love for the Web, engage in conversation, and spread the word to your friends with great Web Foundation content on Facebook, Twitter and beyond.

Latest News

Please let us know if you are planning on giving a talk or have other news relevant to the MW4D IG so that we can add it to the teleconference archives.

- * Minutes of January 18 teleconference published – 15 February 2010
- The MW4D IG held its 32nd teleconference on January 18 2010. The approved minutes are available at <http://www.w3.org/2010/01/18-mw4d-minutes.html>. Previous meeting minutes are available from the teleconference archives
- * Minutes of December 21 teleconference published – 18 January 2010
- The MW4D IG held its 31st teleconference on December 21 2009. The approved minutes are available at <http://www.w3.org/2009/12/21-mw4d-minutes.html>. Previous meeting minutes are available from the teleconference archives
- * Minutes of December 7 teleconference published – 21 December 2009
- The MW4D IG held its 30th teleconference on December 7 2009. The approved minutes are available at <http://www.w3.org/2009/12/07-mw4d-minutes.html>. Previous meeting minutes are available from the teleconference archives

Syndicate this page :: [Next Page >>](#)

Documents & Ressources

The MW4D IG has published the following documents:

- 17 November 2009: Mobile Web for Social Development Roadmap (HTML version)
- See also the PDF (color) version of the roadmap
- See also the PDF (black and white) version of the roadmap

The MW4D plans to develop a set of documents which are described in the Deliverables section of the charter.

The draft version of these notes are developed collaboratively on the dedicated wiki section. The following documents are available:

- MW4D Framework Document
- MW4D Focus Diagram

There is also a set of ressources relevant to the activity of the MW4D and which have been developed by W3C since 2006:

- Executive Summary of the Workshop on the Role of Mobile Technology in Fostering Social Development - Africa region
- Executive Summary of the Workshop on the Role of Mobile Technology in Fostering Social Development
- Executive Summary of the Workshop on the Mobile Web in Developing Countries
- a White Paper on Mobile Web for Social Development
- The previous home page of this work with ressources

gold sponsors

[Read more](#)

DWF Closing Event - Thanks for your participation!

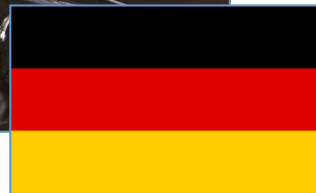
The DWF partners wish to thank all those who made this event a successful one: the [speakers and panelists](#), the [sponsors](#), and all participants. The presentations of the day are linked from the [agenda](#).

Thanks again!

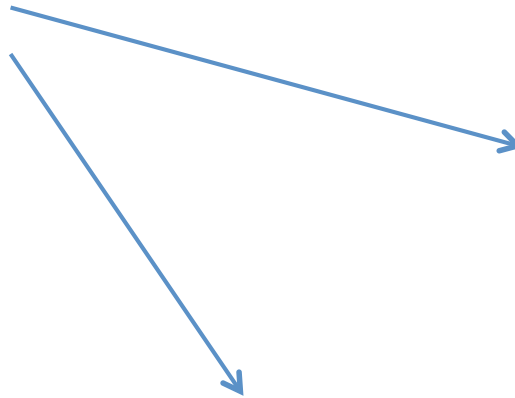
Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

Localization (L10N): Give users what they need



Internationalization (I18N): Prepare yourself for Localization



Proper I18N and L10N: Required for successful Globalization



Traditional topics of I18N on the Web

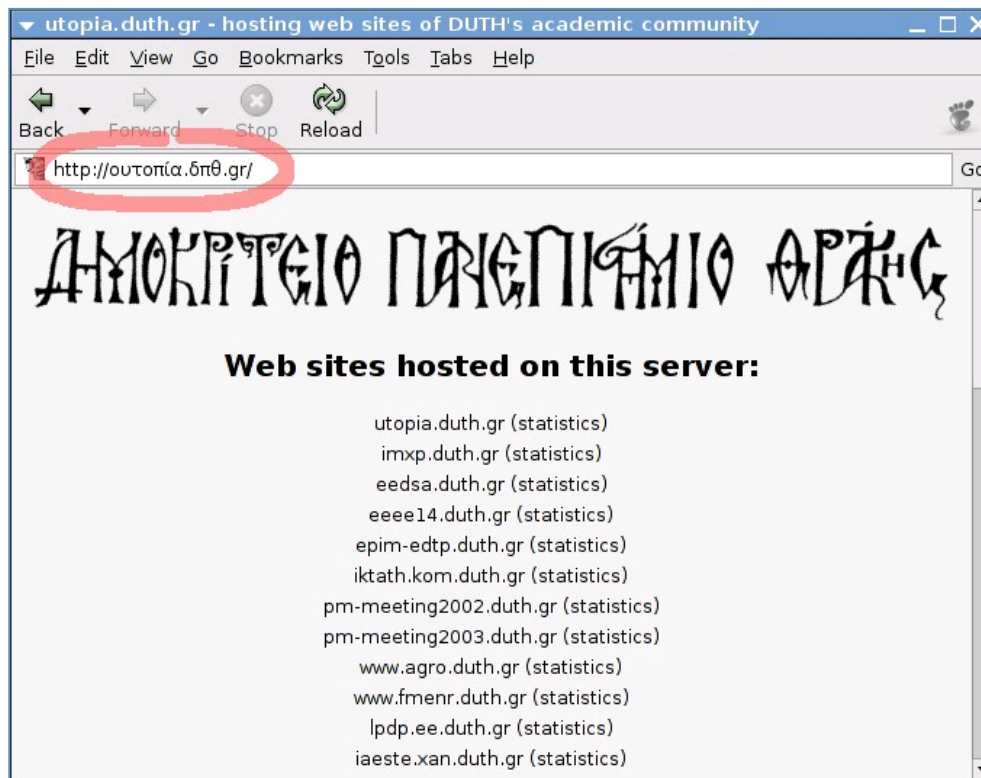
Traditional topics of I18N on the Web

- Use of Unicode in web technologies



Traditional topics of I18N on the Web

Internationalized Domain Names (IDN).



Traditional topics of I18N on the Web

Internationalized Domain Names (IDN). “Actually” possible since 2003, but widely announced last year

The screenshot shows a web browser window with the address bar containing the URL <http://ουτοπία.δηθ.gr/>, which is circled in red. The browser's address bar also shows the text "utopia.duth.gr - hosting web sites of DU".

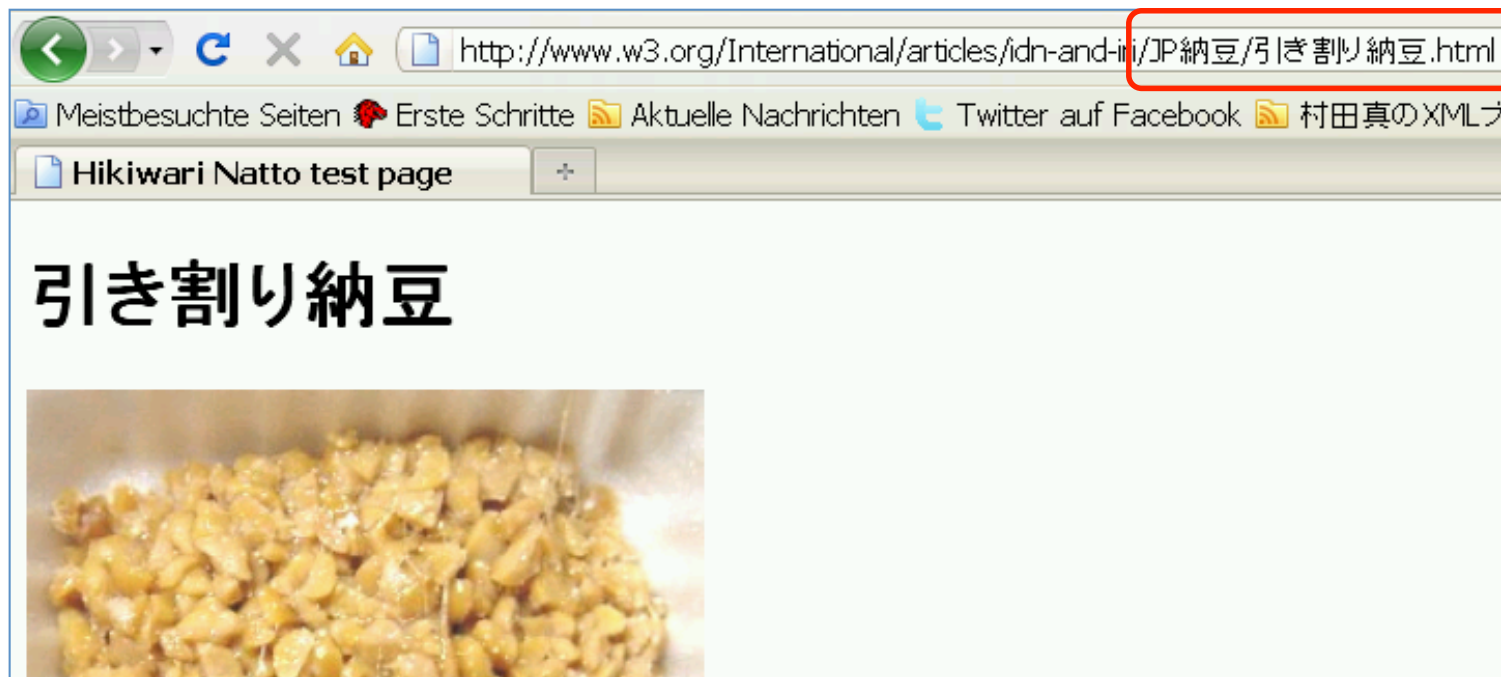
The main content area displays a news article from tagesschau.de. The article title is "Internet-Adressen ohne lateinische Schriftzeichen" and the sub-headline is "Doppeltes Domain-Dolmetschen". The article text discusses the upcoming changes to the Internet-Verwaltung ICANN, allowing for the registration of domain names in non-Latin scripts such as Chinese, Korean, Cyrillic, and Arabic. It mentions that this is a "fantastisch komplizierte Angelegenheit" (fantastically complicated matter) and that the introduction of such domain names is a significant technical change for the Internet.

The article also includes a photo of an internet café in Peking, where many people are using computers. A caption below the photo states: "Internetcafé in Peking: Mehr als die Hälfte aller Internetnutzer benutzt eigentlich eine andere Schrift als die lateinische." (Internet café in Peking: More than half of all internet users actually use a different script than the Latin one.)

The browser's address bar also shows the text "utopia.duth.gr - hosting web sites of DU".

Traditional topics of I18N on the Web

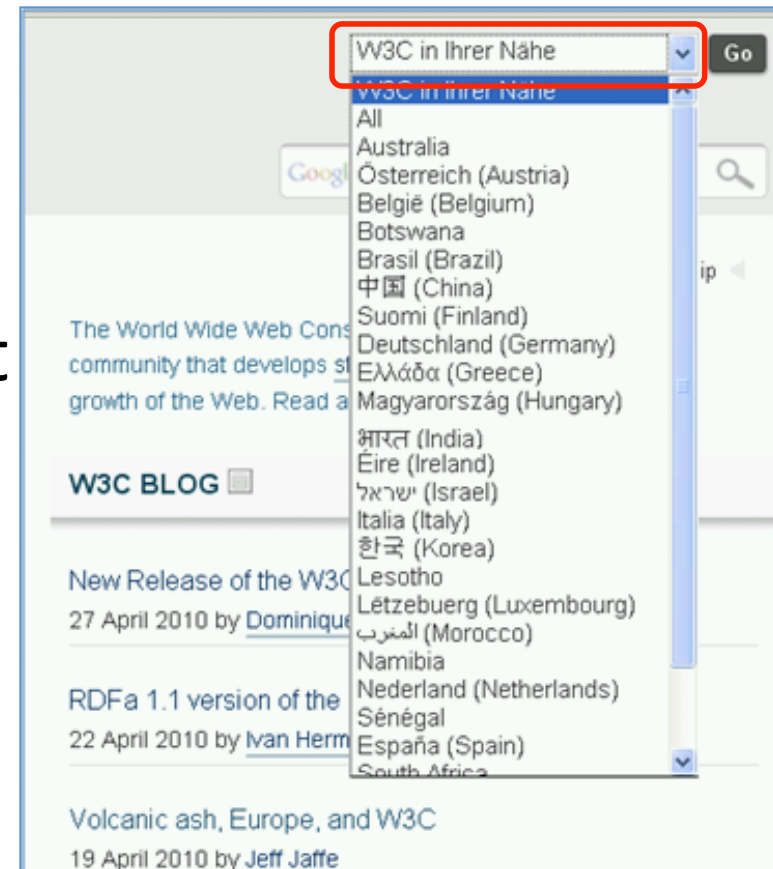
- Internationalized Resource Identifier (IRI)
- I18N in the path of a Web Address, e.g.



Traditional topics of I18N on the Web

- Language tags like “en”, “en-us”, “de”, “ja”, ...

Used e.g. for
content-negotiation:
Give users what
(their browser says) they want

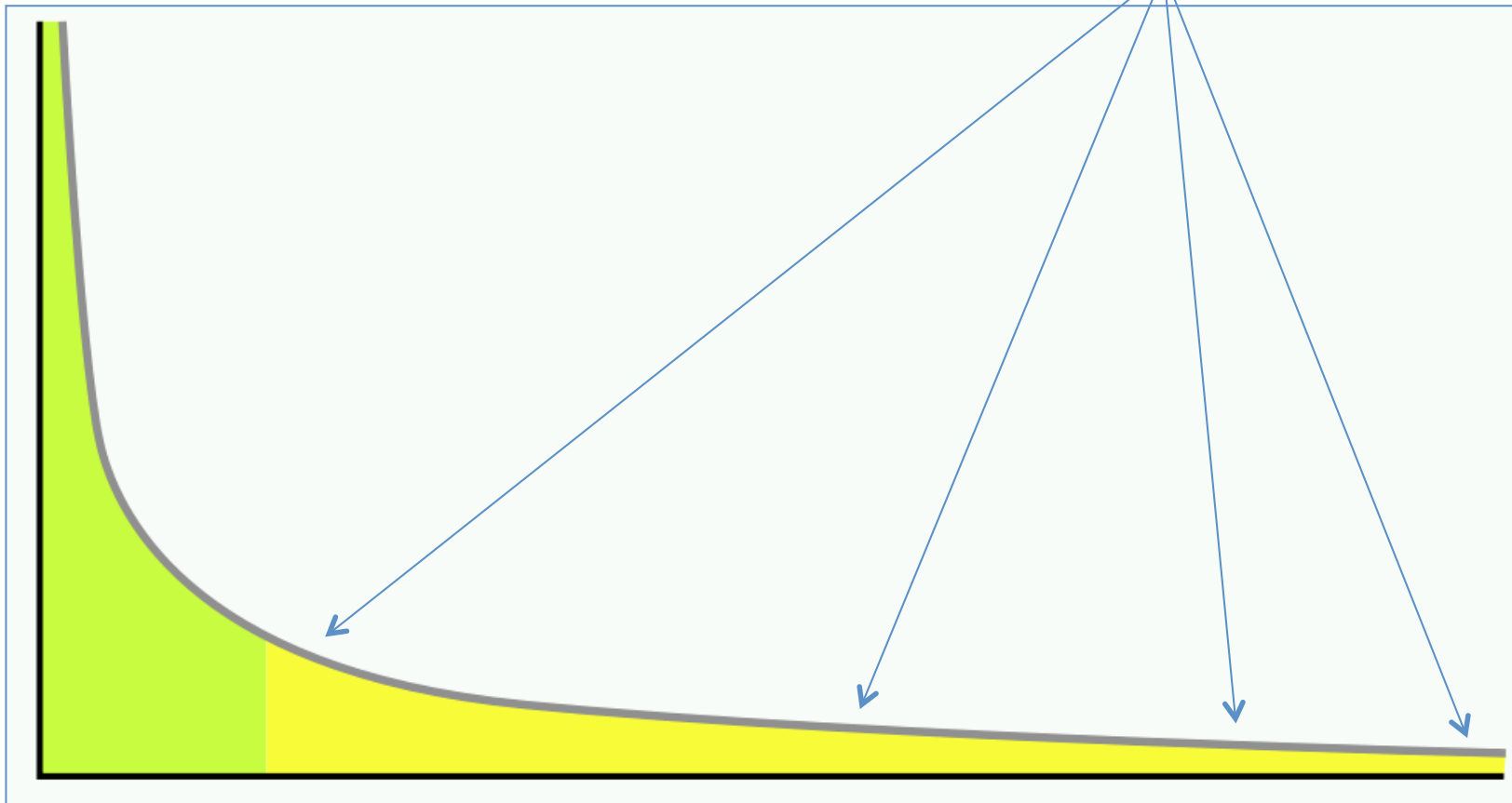


Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

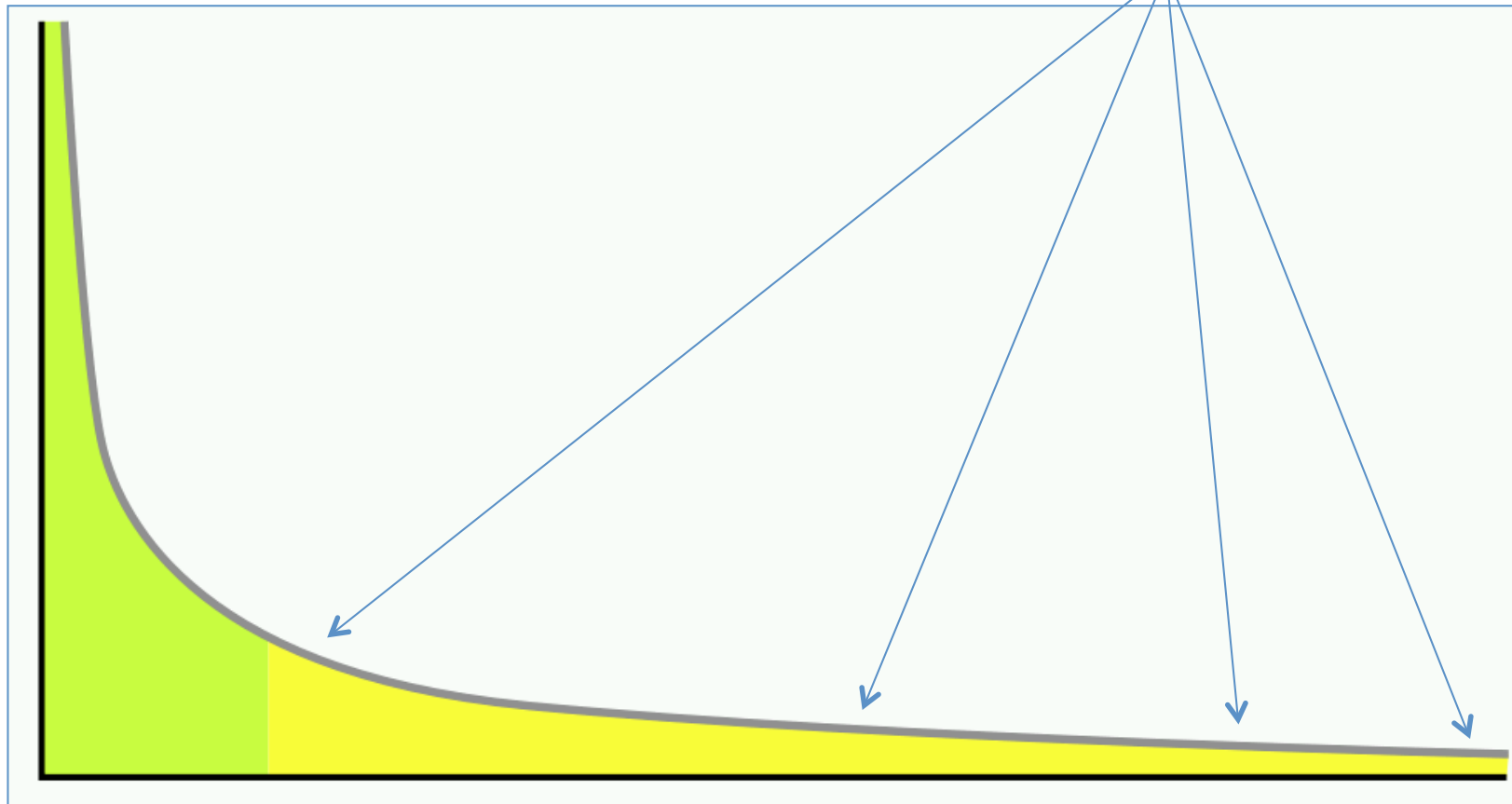
What is the long tail?

- Making money by selling small, but many products



What is the long tail?

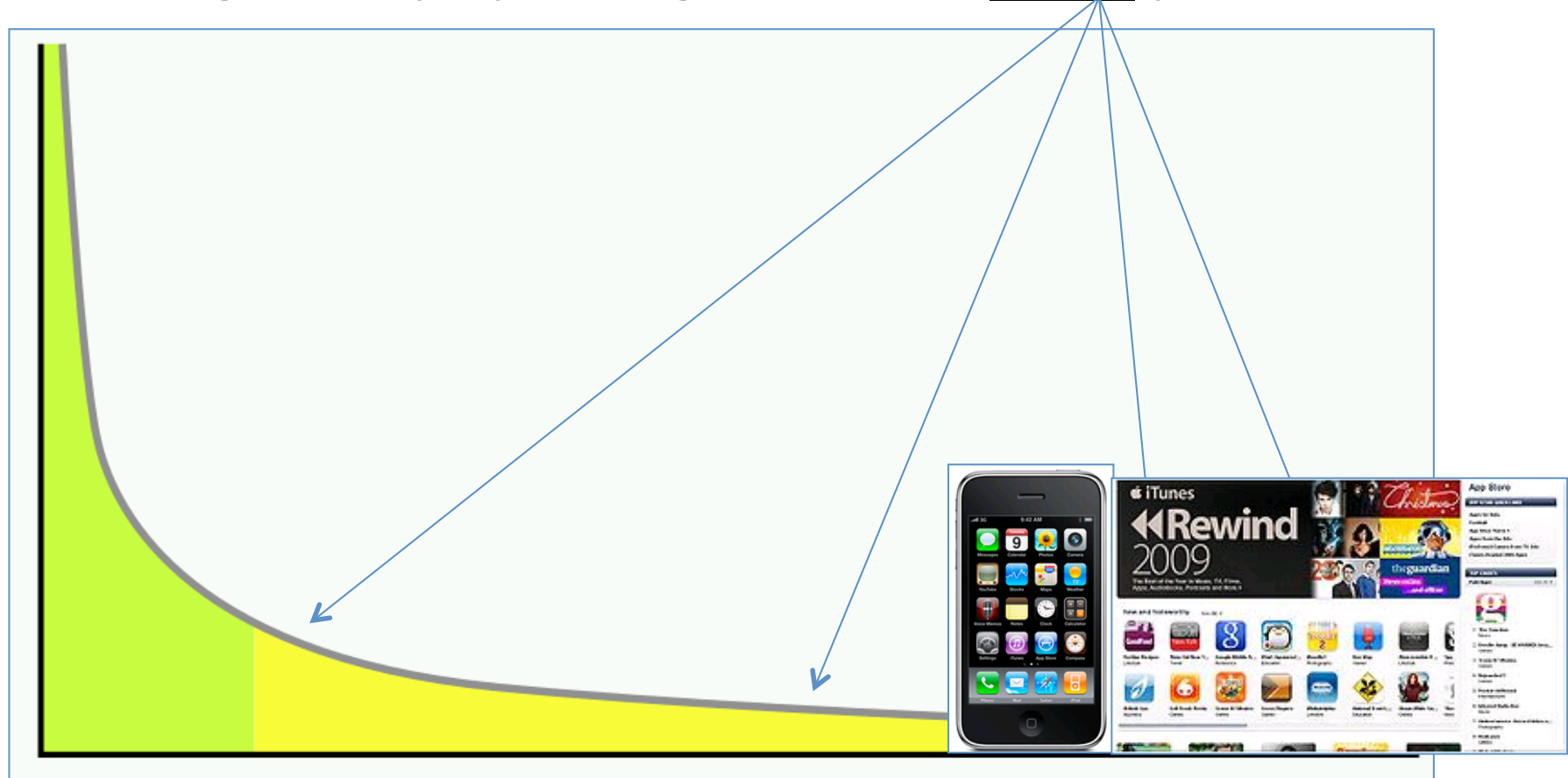
- Making money by selling small, but many products



Example Amazon: "We sold more books today that didn't sell at all yesterday than we sold today of all the books that did sell yesterday."

What is the long tail?

- Making money by selling small, but many products



Example iPhone applications: Third-party applications like games, reference, GPS navigation, social networking, advertising for television shows, etc.

The long tail in I18N / I10N

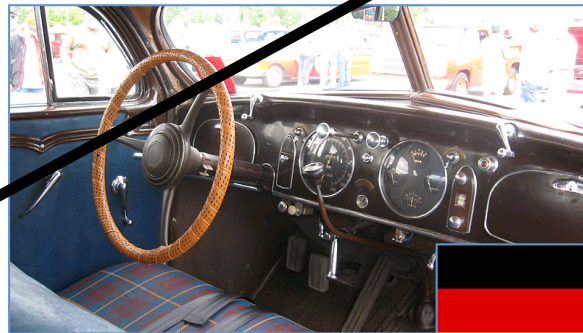
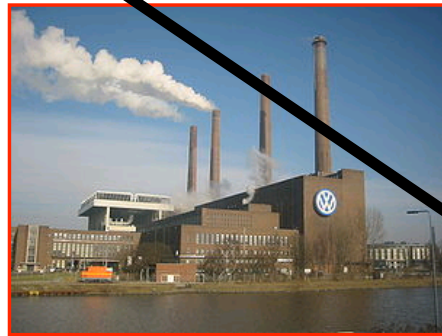
- More specific content and services for many, many devices and users

The collage consists of several overlapping screenshots:

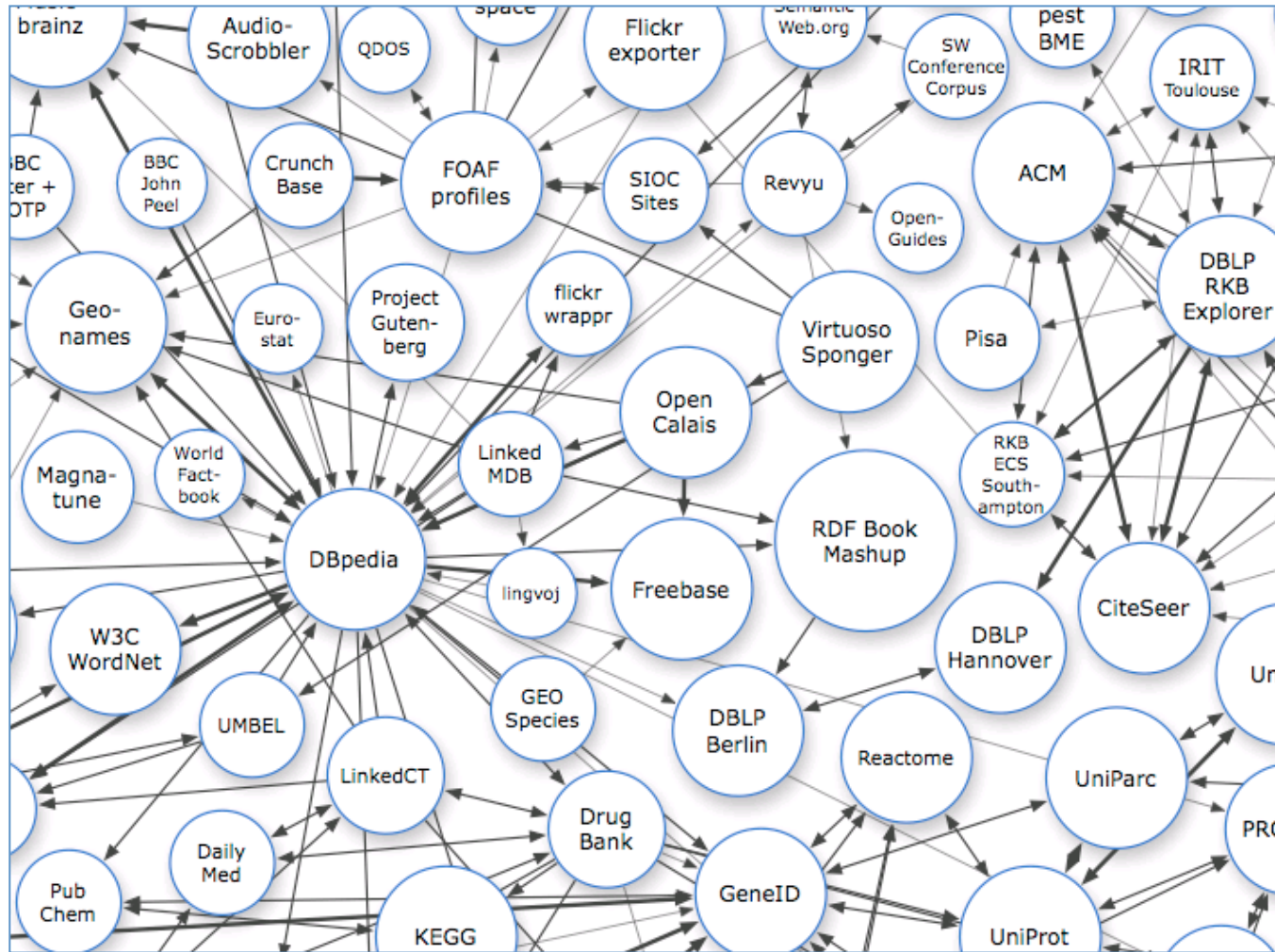
- W3C Korean Site:** Shows the W3C website in Korean, including a navigation menu and a section for CSS Validation Service.
- W3C English Site:** Shows the W3C website in English, featuring a search bar and a section for CSS Validation Service.
- MW4D Page:** Titled "Mobile Web for Social Development (MW4D)", it contains a "Latest News" section with several entries about teleconferences, a "Documents & Resources" section, and a "Current Gold Sponsors" list.
- World Wide Web Foundation Page:** Features a header with navigation links, a main content area with a large image of children, and a "Donate" button.
- Accessibility Page:** Discusses web accessibility, mentioning "Today's Web: Largely Excludes 1-2 Billion People" and "People with disabilities, aging, reading poorly-supported languages, with low literacy and/or lack of technology experience."

The long tail in I18N / I10N

- More specific content and services for many, many devices and users
- Centralized I18N / L10N is too expensive for this scenario ...

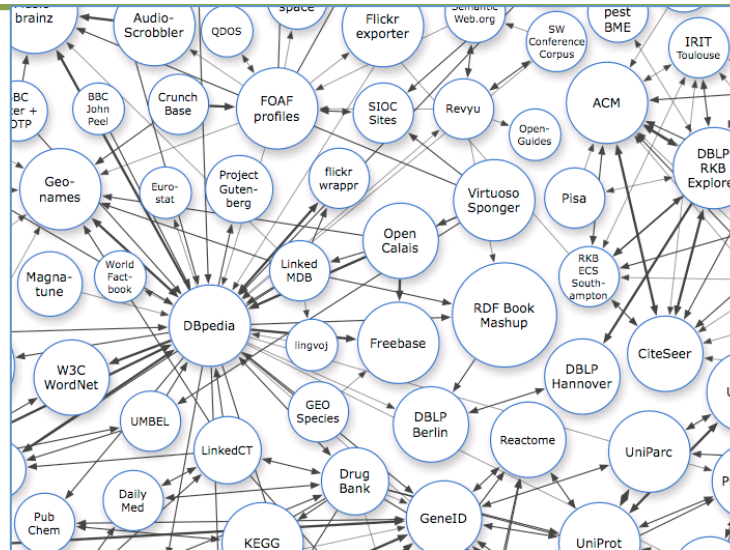


A path to a solution: have a look at how linked data works!



Reaching the long tail in I18N / I10N

- “Linked data” shows a path how to reach the “long tail” market
- Give people a means to create the synergies needed for the long tail, via enhanced, standardized technologies!



Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- The project “Multilingual Web”

Relatively new possibilities of language tags: what does this mean?

- de-1901
- de-1996
- ja-latn
- ja-latn-hepburn
- ja-latn-hepburn-heploc
- ase
- sgn

Relatively new possibilities of language tags: what does this mean?

- de-1901: German, traditional spelling
- de-1996: German, reformed spelling
- ja-latn: Japanese in Latin script
- ja-latn-hepburn: Hepburn spelling
- ja-latn-hepburn-heploc: Library of Congress method
- ase: American Sign Language
- sgn: Sign Language

Who needs a language tag like ja-latn-hepburn-heploc?

- Imagine a (small, but well sold) web application for
 - Automatic and / or manual subtitling Japanese videos on YouTube
 - Making them available for Japanese language learners
- The possibility of creating this cheaply is strongly connected to progress in other area of Web technologies, e.g. HTML5 video accessibility

Application in HTML5 <video> element (still in draft status)

```
<video src="http://example.com/video.ogv" ...>  
<text role="SUB" lang="ja-latn-hepburn-heploc"  
  type="application/smil" src="japanese-lhh.smil"/>  
<text role="SUB" lang="ja" type="text/x-srt"  
  src="translation_webservice/ja/caption.srt"/>  
</video>
```

For details see

<http://www.w3.org/html/wg/wiki/MultimediaAccessibility>

Language tags leads to layout ...

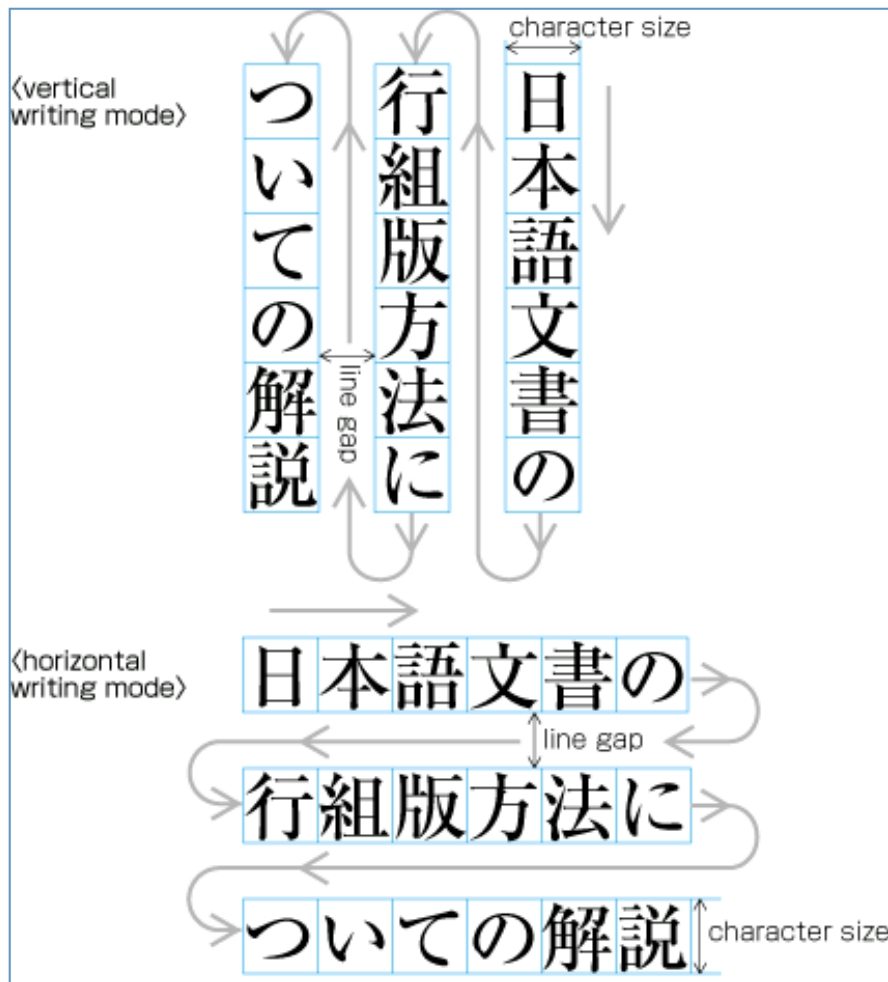
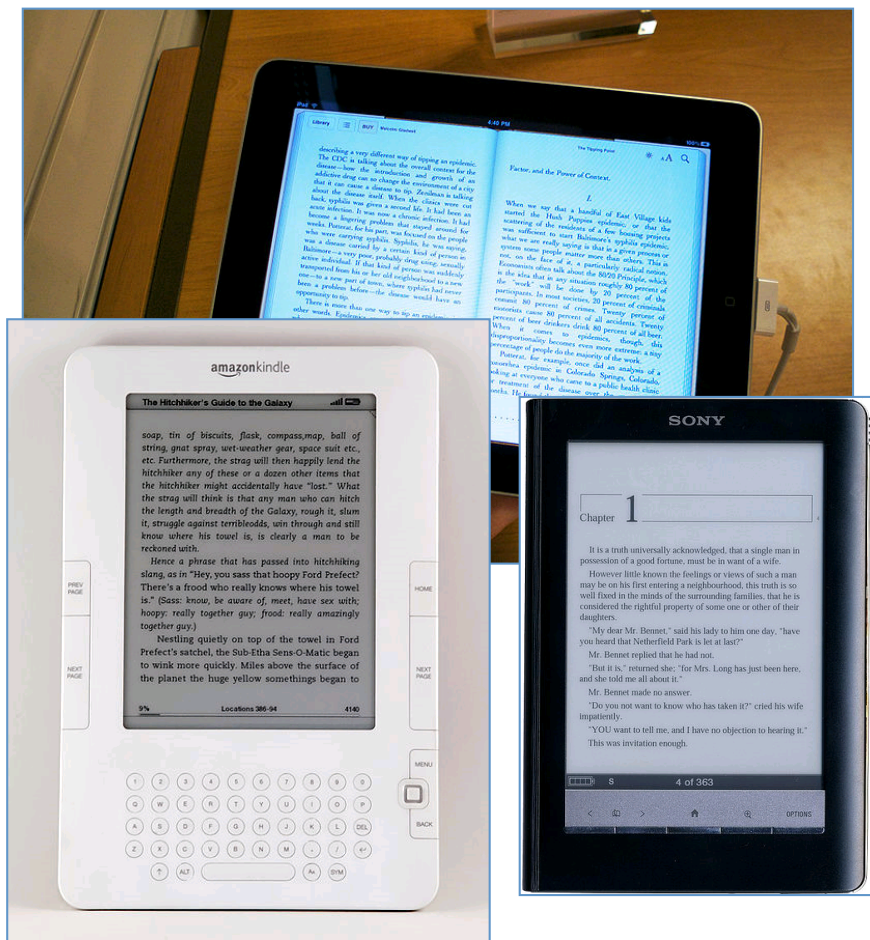
- Already possible for a while: selection of culture-specific glyph images based on language tags

```
<span xml:lang="zh-CN">[雪 zh-CN]</span>  
<span xml:lang="ja">[雪 ja]</span>  
<span xml:lang="ko">[雪 ko]</span>
```



[雪 zh-CN] [雪 ja] [雪 ko]

With new markets – reading electronic books in “Japanese Layout”



Details on Japanese Layout: see <http://www.w3.org/TR/jlreq/>

The tricky bits ...

- An Example: Vertical layout = the same content, just set horizontally? No!
- E.g. marks for “emphasis”: different characters depending on horizontal vs. vertical writing mode

emphasis dots
この表は、法令・公用文書・新聞・雑誌および一般社会で、使用する漢字の範囲を示したものである。（当用漢字表・まえがき）
法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安を示すものである。（常用漢字表・前書き）

emphasis dots
この表は、法令・公用文書・新聞・雑誌および一般社会で、使用する漢字の範囲を示したものである。（当用漢字表・まえがき）
法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安を示すものである。（常用漢字表・前書き）

Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

Bridging technology gaps

- Technologies in Web industry
 - HTML, XML, HTTP, ...
- Technologies in the localization industry
 - Example TMX: Translation Memory Exchange
 - ...
- How should these be combined for fast & cheap localization of content and services for many devices?

Metadata for the rescue!

Metadata 1: XLIFF

- XML Localization Interchange File Format
- Open OASIS Standard for representation of
 - Content to be localized
 - Metadata about the localization process

Basic Example: XLIFF File

- XLIFF files keep source and target content together

```
<trans-unit id="n1">  
  <source>This is a sentence.</source>  
  <target xml:lang="fr">Translation of "This is a sentence."  
</target>  
</trans-unit>
```

Basic Example: XLIFF File

- Metadata in XLIFF files can help to integrate different means of translation (e.g. human vs. machine)

```
<trans-unit id="n1">  
  <source>This is a sentence.</source>  
  <target xml:lang="fr">Translation of "This is a sentence."</target>  
</trans-unit>
```

```
<alt-trans match-quality="100%" tool="TM_System">
```

```
  <source>This is a sentence.</source>  
  <target xml:lang="fr">TM match for "This is a sentence."</target>  
</alt-trans>
```


Metadata 2: Internationalization Tag Set (ITS) 1.0

- W3C-Standard for Internationalization and localization of XML
 - Widely used (HTML, DocBook, DITA, ...) or special purpose formats
 - Content authored in these formats
 - An entry point to the localization tool chain
 - Adopted in localization industry tools like SDL Trados

Basic Principles of ITS 1.0

- Say important things
- About specific content
- In a standard way

Basic Principles of ITS 1.0

- Say important things: Do not translate
- About specific content: all “uitext” elements
- In a standard way: `its:translate="no"`

Two approaches for expressing the same information

Local Approach

<para>

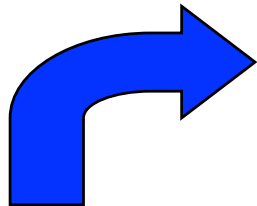
Press the

<uitext its:translate="no">START</uitext>

button to sound the horn. The

<uitext its:translate="no">MAKE-READY/ RUN</uitext>

indicator flashes.



<para>

Press the

<uitext>START</uitext>

button to sound the horn. The

<uitext>MAKE-READY/ RUN</uitext>

indicator flashes.

</para>

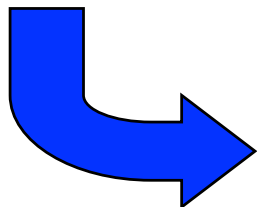
Global Approach

</para>

<its:rules ... its:version="1.0">

<its:translateRule selector="//uitext" translate="no"/>

</its:rules>



ITS 1.0 “Data Categories”

- Translate : Whether the content of an element or attribute should be translated or not
- Terminology: Mark terms and optionally associate them with information, such as definitions
- Directionality: Specify the base writing direction of blocks, embeddings and overrides for the Unicode bidirectional algorithm
- Ruby: Provide a short annotation of an associated base text, particularly useful for East Asian languages
- Language Information: Express the language of a given piece of content
- Localization Note: Communicate notes to localizers about a particular item of content
- Elements Within Text: Identify how an element behaves relative to its surrounding text, e.g. for text segmentation purposes

Why is this important? Example (1) where I18N / I10N Metadata helps in “long tail” localization

Language?	<Собрание версия="1.2-3">
Terminology?	<Объект id="12,,> <НомерОбъекта>OnlineCard</НомерОбъекта>
Codes?	<ВНаличии>123</ВНаличии> <Описание xml:lang="ja">第二発電機</Описание> </Объект>
Footnotes?	<Объект id="64"> <НомерОбъекта>45-7894-456</НомерОбъекта> <ВНаличии>Latest Offer</ВНаличии>
Foreign language expressions?	<Оп xml:lang="ja">手動ウォーター・ポンプ </Оп> </Объект>
Annotations for readers?	</Собрание>

Example (2) where knowledge about I18N markup usage helps in “long tail” localization

Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder.



Kurokami maibutsu gate (腹五社神社黒神埋没鳥居), Sakurajima Island.

Suitable Markup?

```
<image src="kk-torii.jpg" height="180" width="240" caption="Kurokami maibutsu gate (腹五社神社黒神埋没鳥居), Sakurajima Island." />
```

Better:

```
<image src="kk-torii.jpg" height="180" width="240">
  <caption>
    Kurokami maibutsu gate (
    <span xml:lang="ja">腹五社神社黒神埋没鳥居</span>
    ), Sakurajima Island.
  </caption>
</image>
```

Adopted from Richard Ishida (W3C)

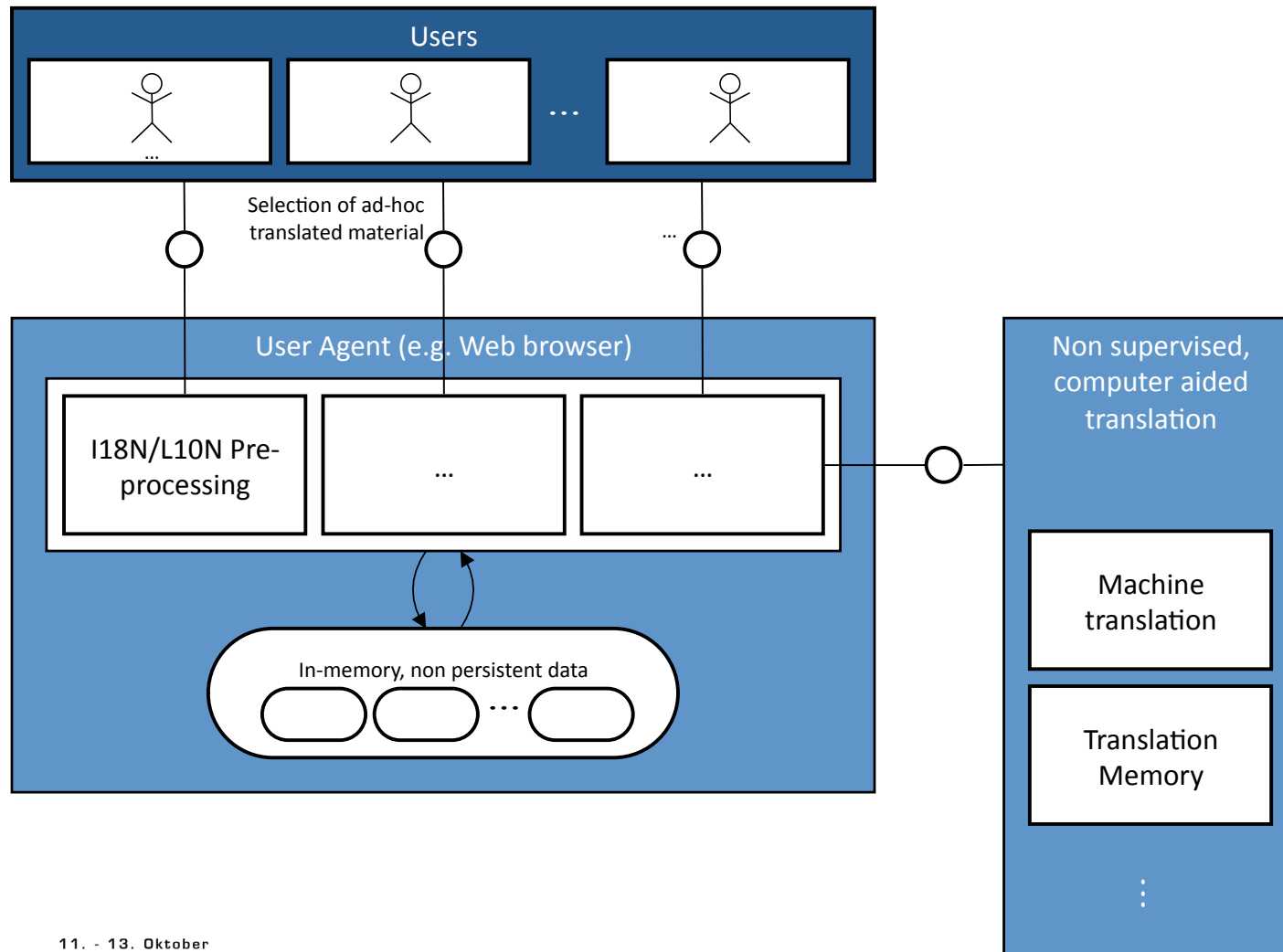
XLIFF and ITS 1.0

- ITS 1.0: entry point in the localization chain
- A pre-requisite for properly internationalized vocabularies and content
- XLIFF: the meat of the localization chain
- ITS 1.0 makes XLIFF creation and processing easier

XLIFF and ITS 1.0 example: ITS2XLIFF

- See <http://fabday.fh-potsdam.de/~sasaki/its/>
- XSLT-based round tripping tool for generation of XLIFF from XML with ITS markup, and integration of translated content into the original XML
- Open source standard technology based - one example of how “long tail” content localization can become easier

Future usage scenarios for ITS 1.0: “Long tail” L10N via the Web



Who can help to make that happen?

- Of course you!
- If content authors & developers don't use ITS, tools which want to generate XLIFF from let's say HTML will produce a lot of crap
- We need you to fix that!

Why we need you for the multilingual web

- Input from www.postbank.de
„Ob Postbank direkt, Online-Banking, Online-Brokerage oder myBHW. Die häufigsten Fragen zu unseren Transaktionssystemen finden Sie an dieser Stelle.“
- Output via Google translate
“Whether Postbank direct, online banking, online brokerage or myBHW. Frequently asked questions about our transaction systems can be found at this location.”

Why we need you for the multilingual web

- Input from www.postbank.de
„Ob Postbank direkt, Online-Banking, Online-Brokerage oder myBHW. Die häufigsten Fragen zu unseren Transaktionssystemen finden Sie an dieser Stelle.“
- Output via Google translate
“Whether Postbank direct, online banking, online brokerage or myBHW. Frequently asked questions about our transaction systems can be found at this location.”

Fixed terminology should not have been translated. If a content author / editor / developer (= you too!) would have used ITS “translate”, Google translate would have worked.

What do we learn from this? Automatic language processing is better ...

- ... if you help it with metadata
 - About what is translatable or not
 - About terminology
 - About segmentation (“What is a footnote?”)
 - ...
- Your employer will like it too
 - (Long tail) localization with high quality metadata gets better
 - The whole localization process gets cheaper 😊

The question is now: Who will start?

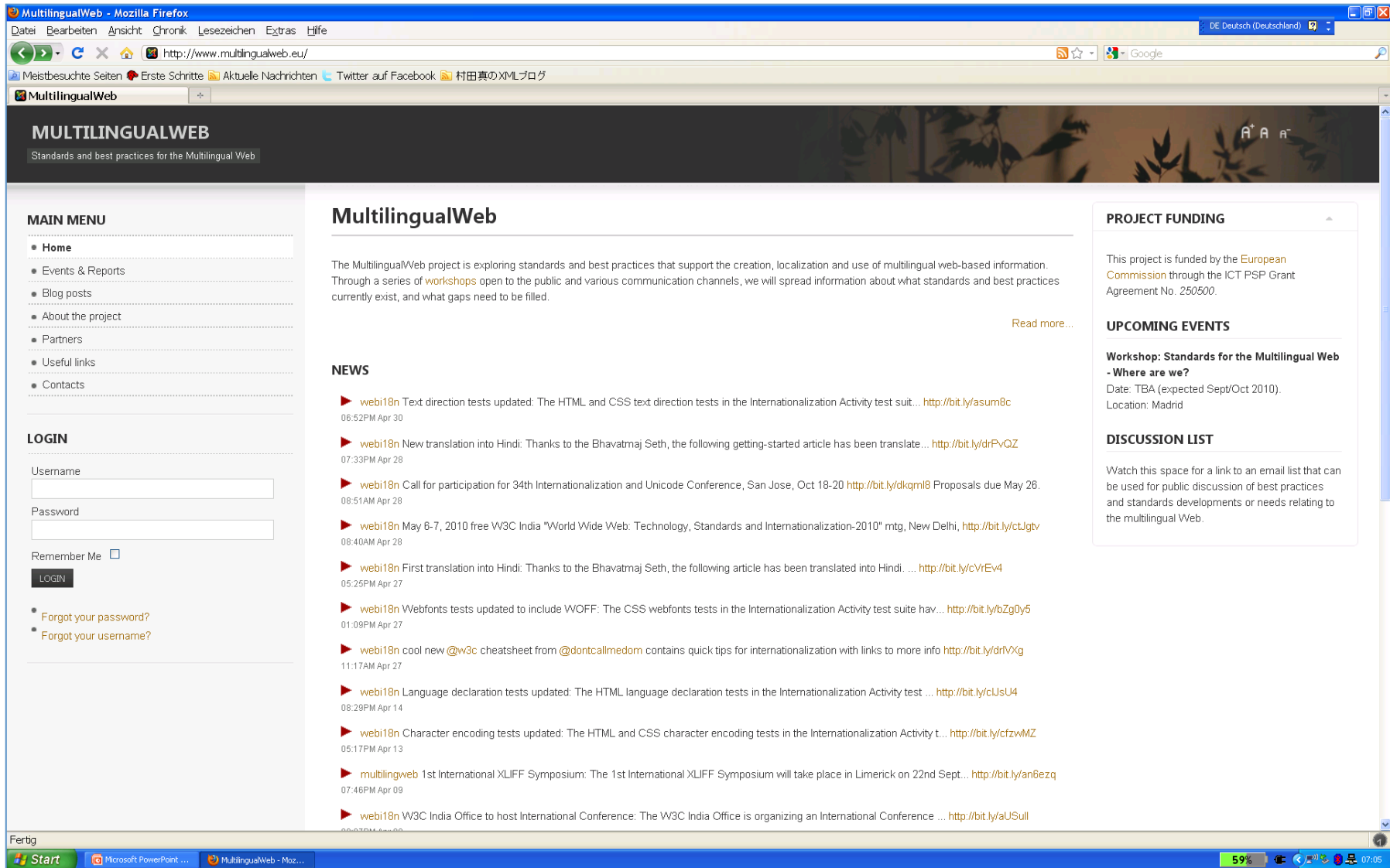
- Content creators don't use metadata since nobody processes it
- Nobody processes metadata since there is none
- The first company who breaks that vicious circle and develops a successfully model of deploying metadata (see previous slides) will get rich 😊 😊 😊

The question is now: Who will start?

- Some people already have started: WAI-ARIA (cf. talk from Tomas Caspers at this conference)
 - Metadata for “roles” of specified elements
 - “I am navigation”
 - “I am a tab”
 - ...
- Used e.g. by screen readers
- Shows common aspect of barrier free web design and web design for the “automatable, multilingual web”: appropriate metadata

Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”



See <http://www.multilingualweb.eu/>

Background

- Participants from industry and academia (e.g. computational linguistics)
- Aim: Bridge the gaps between industries (and research areas) described
- Education about new developments (e.g. in the area of language tags, layout, web based localization) – like in this presentation 😊
- More mutual understanding of users needs and what tools can do
- Support via special purpose tool
 - First example “I18n checker”
<http://rishida.net/tools/i18nchecker/>

Participants from: Web, Localization industry, (automatic) translation research, ...

- ERCIM/W3C: coordination
- CNR-ILC
- Facebook Ireland
- The University of Applied Sciences (UAS) Potsdam
- Institut Josef Stefan (JSI)
- Institutul de Cercetari Pentru Intelgentia Artificiala (RACAI)
- The Language Technology Centre
- Lionbridge Belgium
- Microsoft Ireland
- Opera Software
- SAP
- The Translation Automation User Society (TAUS)
- Teknillinen Korkeakoulu
- University of Oviedo (ILTO)
- Universidad Politécnica de Madrid (UPM)
- The Language Resource Centre
- University of Economics, Prague
- Transware Ltd (WeLocalize)
- XML-INTL

Workshops as a means for community building - topics

- The landscape of multilingual Web standards and best practices
- Authoring of the Multilingual Web
- Translation tool support (with focus on standards like ITS 1.0, XLIFF, TMX)
- Further topics – to be decided
 - Your input is more than welcome!

1st Workshop: “The Multilingual Web – Where Are We?”

- 26-27th October, Madrid
- Goal: Bring developers (you!), content creators, localizers, users, machine processing folks and policy makers together
- Provide input for upcoming work
- Program details at

www.w3.org/International/multilingualweb/madrid/program

META-NET

- EU-funded project, closely related to “Multilingual Web”
- Main aim: build an alliance for improving language technologies in Europe
- Large: soon 40+ participating organizations in 30+ countries
- Very important: bring users of language technology in

META-NET

- Users and language technology companies = in Europe not only large companies, but more and more small SMEs
- Target of META-NET are these small and fast units – including you 😊
- EU has started special funding programs for SMEs – see <http://tinyurl.com/eu-It-sme> (“objective 4.1”)

META-NET

- Event: META-NET Forum
- Brussels, November 17th/18th
- Aim: Bring users / language technology developers / policy makers together
- Discuss a road map for the next 10 years of language technology road map and its applications
- Details and registration at

<http://www.meta-net.eu/events>

Overview

- Multilingual Web – what’s that?
- I18N and L10N – traditional topics
- “The long tail” and its consequences for the multilingual web
- I18N and L10N on the Web - revisited
 - Traditional topics & new aspects
 - New: bridging technology and market gaps
- Projects “Multilingual Web” and “META-NET”

Summary

- The basic infrastructure of the multilingual Web is set
- Nevertheless, even core parts like language identification have are being updated to easy international use for many & diverse audiences
- Localization is getting diverse across industries (Web, Localization, including automatic translation)
- Specific (Meta)data formats like XLIFF and ITS 1.0 can help bridging gaps between players and users in the field
- We need you as a metadata user to realize faster and cheaper localization
- The “Multilingual Web” project aims at being a place for gathering possibilities for a “long tail” multilingual web – a web for really everyone’s business!
- The “META-NET” project aims at building an alliance between language technology providers and users to make the long tail localization model happen

Thank you for your attention!
Vielen Dank für Ihre
Aufmerksamkeit
ありがとうございました