MultilingualWebLT

MultilingualWeb-LT:
Putting the World in the World-Wide Web

Arle Lommel

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

LAIA Standards Initiative

Administered by: W3C®

Funded by:

MultilingualWebLT

SOME BACKGROUND

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

MultilingualWebLT

The Vision of the World Wide Web

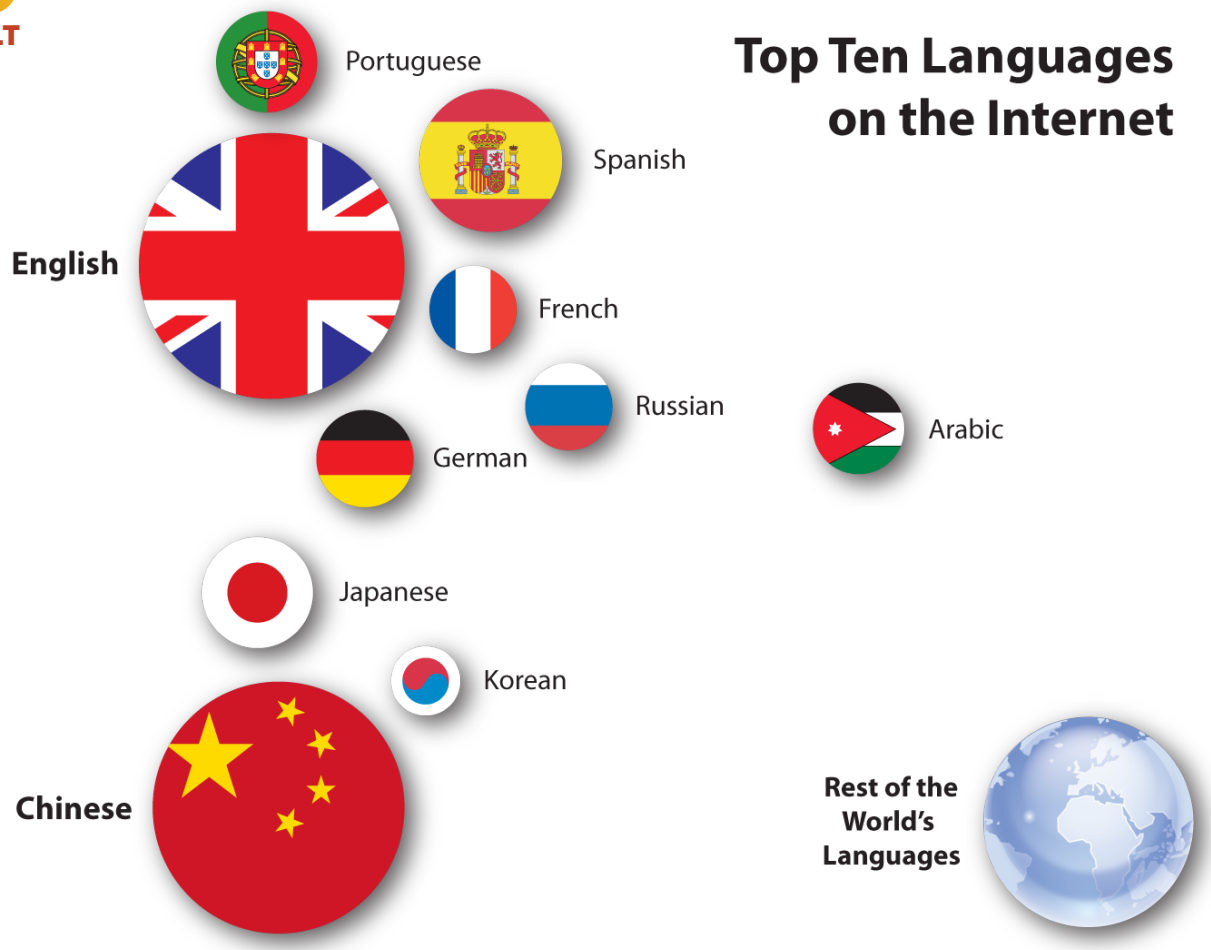
This works at the data level: you can access the Internet from (almost) anywhere in the world

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

The Reality: Multiple Webs



Based on
http://www.internetworldstats.com/stats7.htm

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

Multilingualism Becoming More Important

- Since 2000:
  - Chinese grew ~1500%
  - Arabic grew ~2500%
  - Russian grew ~1800%
  - Top ten languages grew 421%
  - Rest of the world's languages grew 589%
- Critical issue in Europe

Based on
http://www.internetworldstats.com/stats7.htm

Metadata

- To reduce costs, we need to streamline the flow of data
- Two ways
  - File level metadata (Linport)
  - Inline metadata

The MultilingualWeb-LT Project

- Continuation of the MultilingualWeb project (http://www.multilingualweb.eu)
- LT stands for "Language Technology"
- A W3C standards effort
- Funded by the European Commission

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

The MultilingualWeb-LT Project (2)

- Addresses Web (HTML5) content
- Addresses "deep Web" content (i.e., content in databases, CMSes, etc., that will appear on the Web)
- Addresses relation to other standards (e.g., XLIFF)

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

MultillingualWeb Partners

- Cocomore
- Dublin City University
- ENLASO
- German Research Center for Artificial Intelligence (DFKI)
- Institut Jožef Stefan
- Linguaserve
- Lucy Software and Services
- Microsoft
- Moravia
- Trinity College Dublin
- University of Economics, Prague
- University of Limerick
- Vistatec

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

MultilingualWebLT

Why Not Google Translate?

- Quality is an issue
- Lack of integration
- No social interaction

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Standards Initiative GALA

Administered by: W3C®

Funded by:

MultilingualWebLT

Humans Are…

- Very good at:
  - deduction
  - dealing with (implicit) context
  - adding intelligence
  - dealing with exceptions
- Very bad at:
  - doing things quickly
  - doing the same thing over and over

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

MultilingualWebLT

Machines Are…

- Very good at:
  - doing things quickly
  - doing the same thing over and over
- Very bad at:
  - deduction
  - dealing with (implicit) context
  - adding intelligence
  - dealing with exceptions

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

LAV Standards Initiative

Administered by: W3C®

Funded by:

MultilingualWebLT

Intelligence Creates Quality

- To fix quality (for humans or machines), you need intelligence…
- Intelligence allows you to deduce things you would not otherwise be able to understand
- Should "Different Types of Windows Are Very Useful" be?
  - Verschiedene Arten von Windows sind sehr nützlich
  - Verschiedene Arten von Fenster sind sehr nützlich

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

LAIA Standards Initiative

Administered by: W3C®

Funded by:

MultilingualWebLT

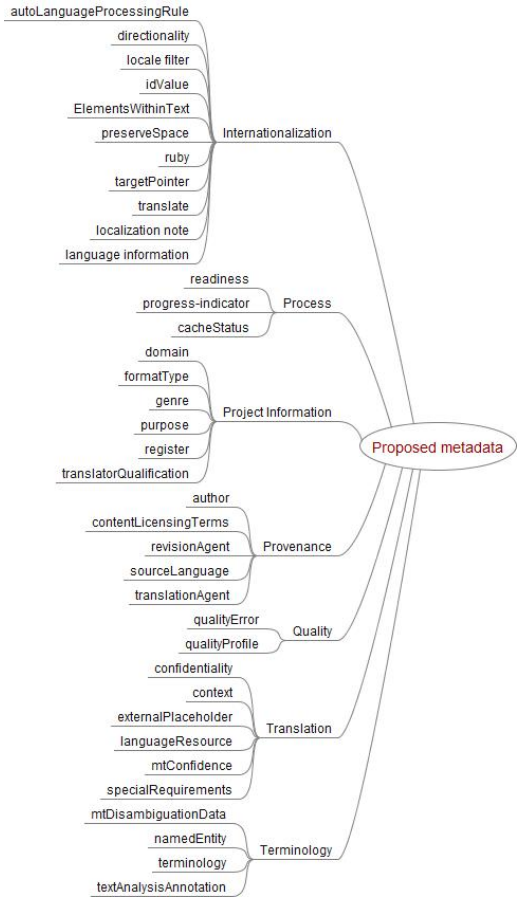
If Machines Cannot Add Intelligence…

- Then we need to make data intelligent
- Apply what humans do well to enable machines to do what they do well

Examples

- The “translate” attribute (already in HTML5):
  - He bought a Super Soaker™ water gun at Joe’s Toy Emporium
- “Domain”:
  - Windows = Windows (IT)
  - Windows = Fenster (Construction)

Proposed Scope



Intended Outcome

- Lower barriers to information access
- Lower cost of translation (for human translation)
- Increase quality (for all translation)
- Enable linking of data across languages (improve quality for all)

Options to Get Involved

- MultilingualWeb homepage: http://www.multilingualweb.eu/
- Join the MultilingualWeb Workshop http://www.multilingualweb.eu/documents/dublin-workshop/dublin-cfp
- Contribute to the specification
- Join the Working Group
- Contact arle.lommel@dfki.de for more info

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by:

MultilingualWebLT

ANY QUESTIONS?

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Administered by:

Funded by: