

"From the Lab to the Market" Commercialising MT Research John Tinsley Director / Co-Founder

Beregovaya et al. (Eds.) Proceedings of AMTA 2014, vol. 2: MT Users Vancouver, BC © The Authors

AMTA. 24th Oct 2014. Vancouver

105

What is Iconic Translation Machines?

We provide Machine Translation solutions with Subject Matter Expertise

HQ: Dublin, Ireland Founded: 2012





Part 1: The journey from the lab to the market

Part 2: The technology that took the journey



Innovation in academia

Typically Cutting edge techniques not seeing the light of day

More Recently Basic research → Applied research

New Performance Metrics Industry collaboration, software licenses, spin-outs

Lab to Market: The Starting Point... The Lab

The Lab Dublin City University

The Funding European Union (FP7 PSP)

The Goal Adapt existing technology for patent machine translation

So what now?





Lab to Market: Technical Development

The Process Build with working group: release early, release often

Engagement Identify broader user base (patent professionals, translators) and field test

The Outcome Well-developed, well-performing prototype with a user base

Is there a business in this?









• Investment





Part 1: The journey from the lab to the market

Part 2: The technology that took the journey



We provide Machine Translation solutions with Subject Matter Expertise



An "ensemble" MT architecture built with Linguistic Engineering

What is Linguistic Engineering? Data Engineering







The Challenge of Patents

Long Sentences

US 7,752,3

network connections and preferably but not necessarily preemptive multitasking and multithreading. The interface of the present invention as it appears to the outside world (i.e., programmers and developers who provide access to end users and programmers who provide Content Access Modules to 5 Content Engines used by content publishers) is independent of both the operating systems and the programming languages used. Adapters can be built allowing the tiers of the system to cooperate even if they use a different operating system or a different programming language. The protocol of 10 the present invention can be implemented on top of networking standards such as TCPI P. It can also take advantage of inter-object communication standards such as CORBA and DCOM. The object model of the news univention can be

requesting cheen of the protonty of chemis, thus provides an input field and that transmits to the server within the user session a plurality of queries to retrieve content from the server, wherein each of the plurality of queries are consecutive and form a lengthening query string for retrieving content from the server, and wherein each subsequent one of the plurality of queries extends the lengthening query string by one or more additional charnetters; and

server, which is configured to access one or more content sources that store content and that can be accessed by the server to respond to the queries from the clients, wherein the server further includes a unified query and result cache common to the obuestive of clients and that stores

Longest Sentence: 1,417 words

Largest single document: 249,322 words

maximum stress of 1.2 to 3.5 N/mm<2> and a maximum elongation of 700 to 1,300% at 0[deg.] C.

Beregovaya et al. (Eds.) Proceedings of AMTA 2014, vol. 2: MT Users Vancouver, BC © The Authors

The foregoing description of the present invention has been provided for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the species forms direlowed. Many modifications and variations 5° will be apparent to the prescriptioner skilled in the art. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application, thereby enabling others skilled in the art to understand the invention for various embodiments and with various 5 modifications that are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the following claims and their equivalence.

What is claimed is:

- A system for searching at a client for content at a server ir other content sources, comprising:
- a communication protocol that provides in asynchronous connection between each of a plurality of clients and a server, and allows each client to send, as part of a user a session, a plurality of consecutive query strings to query the server for content;

queries are a plurality of additional characters to be added to the increasingly focused query string.

5. The system of claim 1 further comprising a server repository for storing content information and for use as a dynamically updated query and result cache in returning increasingly relevant content to the client from the server repository in response to automatically matching the increasingly focused query string, prior to retrieving matching content from the content sources if the relevant content was not found in the server repository.

6. The system of claim 1 wherein the system is further configured to access a pluratily of content sources via a content engine and content channel associated with each content source, and wherein the server comprises a plurality of query and result caches, including a query cache associated with each particular content source that stores previously determined results from that periodular content source.

7. The system of claim 1 wherein only the difference between a claim's current data set and the cleim's requested data set is transmitted over the network, and wherein the server only returns those results that were not sent in a previous results message for the same query.



What is Linguistic Engineering? Data Engineering





An "ensemble" architecture

Data Engineering + Linguistic Engineering



What is the value for users?

De-risking the machine translation proposition



Our Prerequisites

- + No data needed
- + Systems are ready to go
- No upfront cost
- Evaluate immediately

Customisation. Refinement.

» Incorporation of user feedback
» Incremental training with post-edits
» Tuning for specific input types



Case Studies

- 1. What this approach means straight up in terms of quality...
- 2. Productivity gains from using these systems...
- 3. As a foundation for client customization...





Portuguese to English



Translation Machines



German to English







Business Need

Machine Translation technology for the legal industry

Iconic had a domain-specific MT solution for that industry





Process

Translation samples required for initial evaluation

Delivered immediately and initial results were positive





Performance

"MT delivered measurable productivity gains from the outset"

>20% productivity increase for translator post-editing Iconic output





Chinese to English

- Modify our patent machine translation engines for "Written Opinions" on patents
- 0.25% new data, 2 new ensemble processes







Essentially out of domain – not viable for post-editing







After customization – 25% gain in productivity





Take home messages...

- Do you have technology that can solve a problem?
 Validate this!
- Is there a market for this technology?
 - Find product/market fit!
- Go for it!
- This is what we did in developing domain-adapted MT solutions with subject matter expertise for LSPs and data providers.
- Enjoy the ride!







john@iptranslator.com





Iconic Translation Machines



