

A designerly approach to interactive cross-language information retrieval

Daniela Petrelli

Art & Design Research Centre

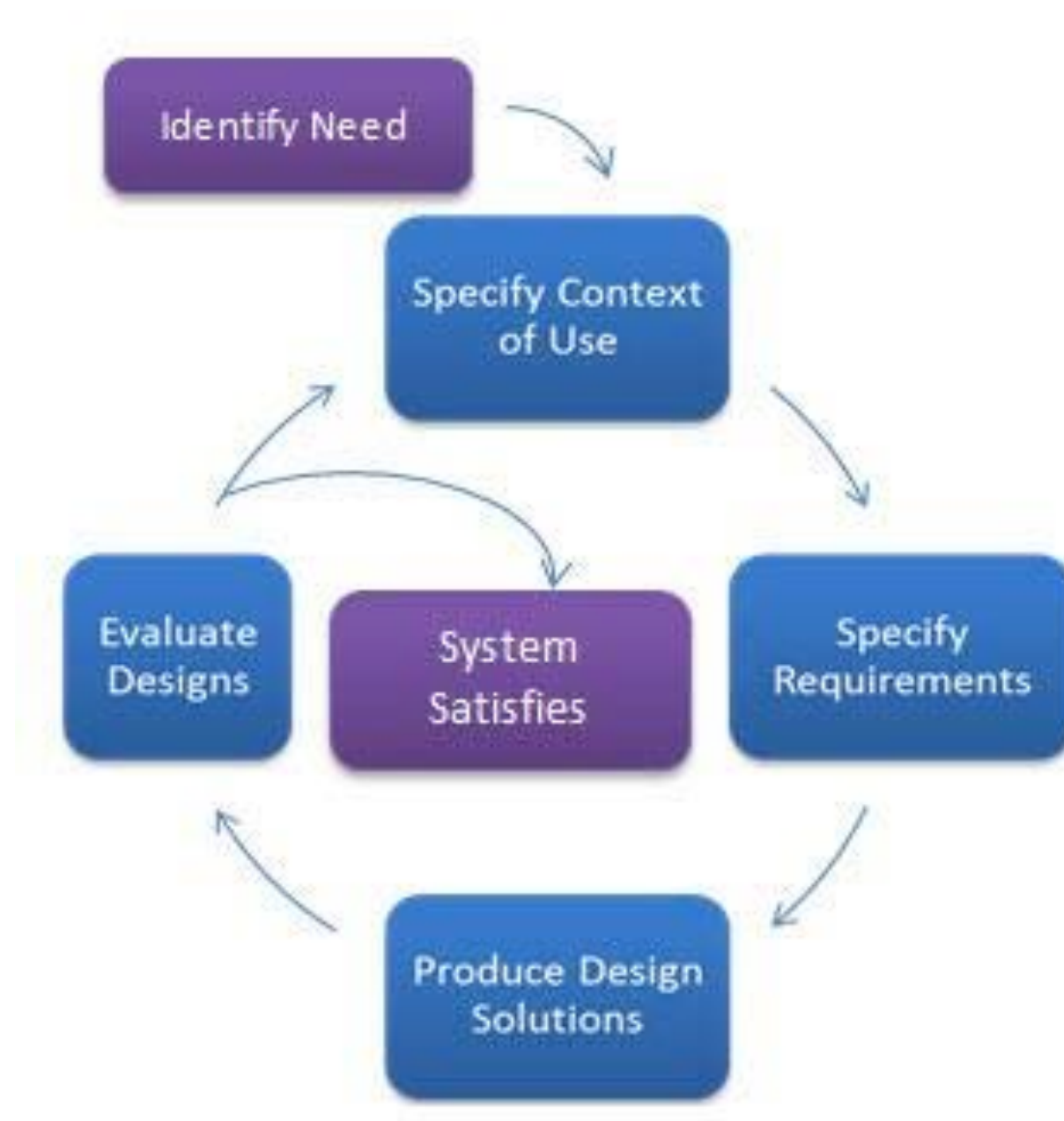
Sheffield Hallam University

UK (still EU!)

d.petrelli@shu.ac.uk

@DanielaPetrelli1

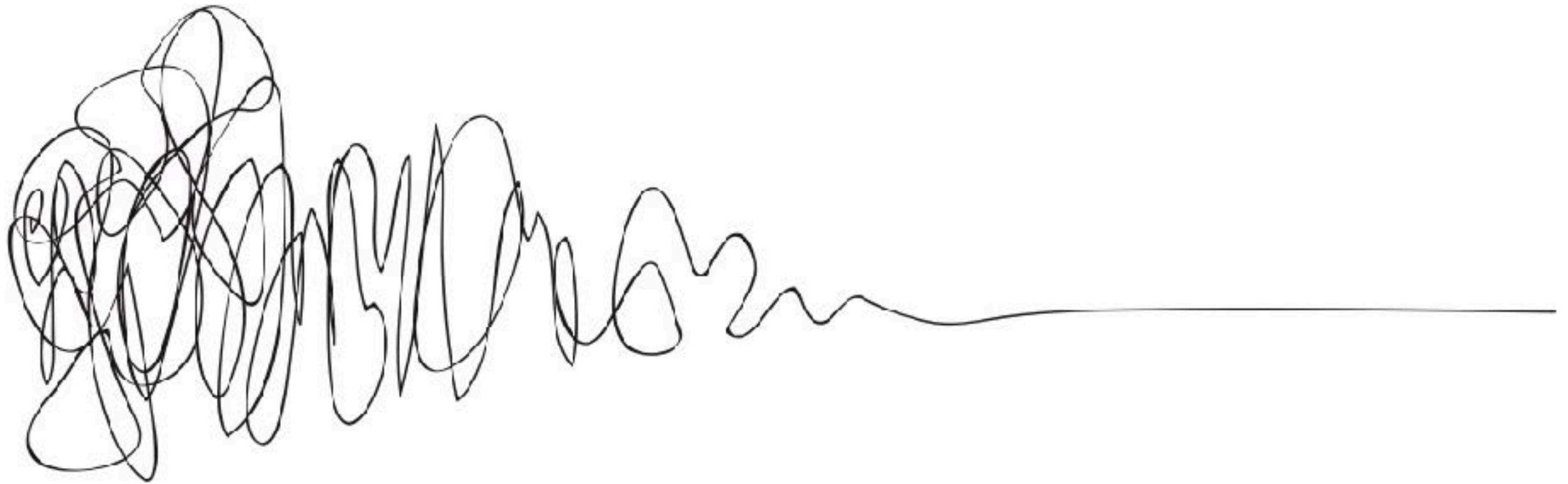
The User-Centred Design Process



The Design Process

uncertainty / insight

focus / implementation



exploration

concept / prototype

product

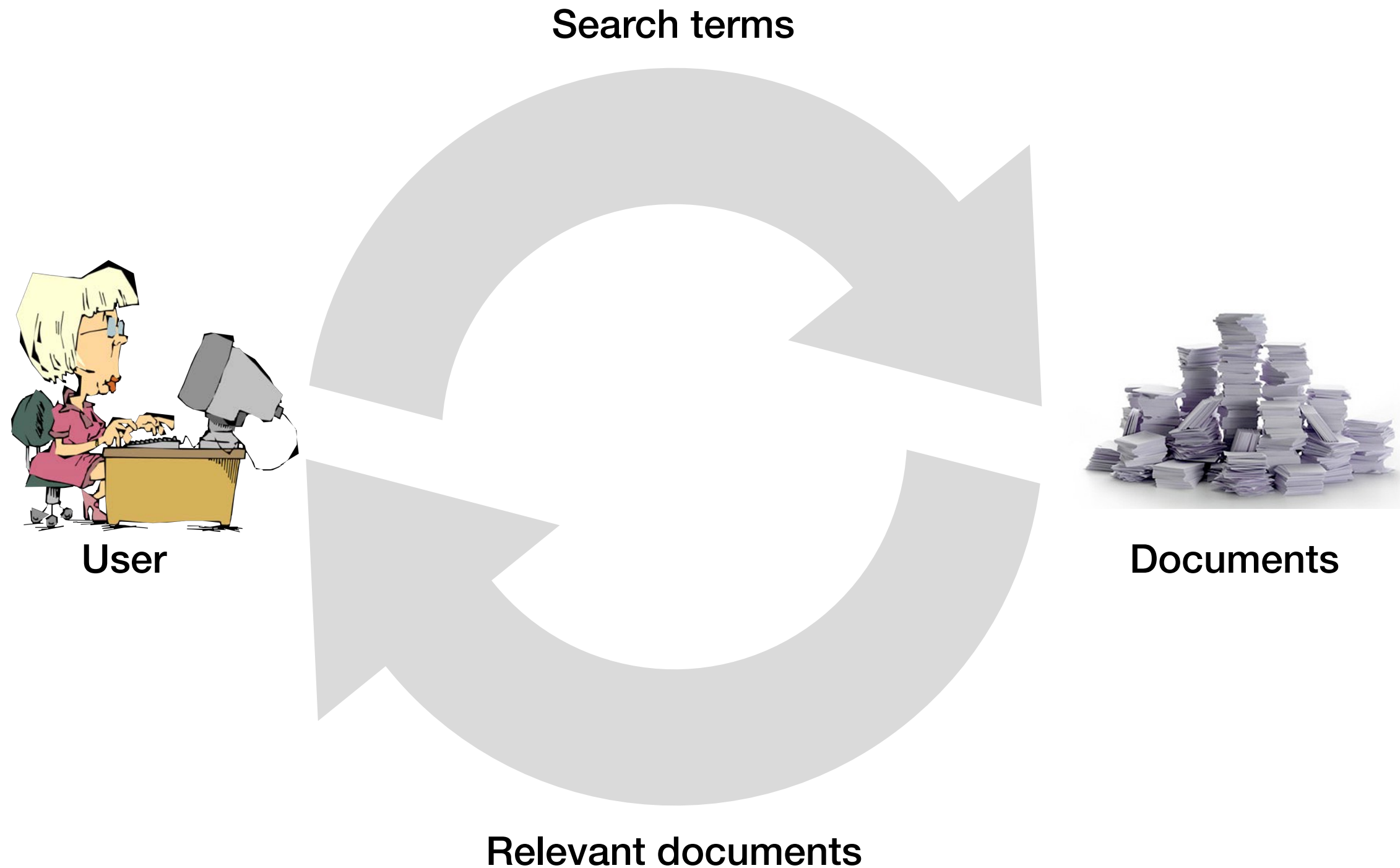
Cross-language information retrieval (CLIR)

the retrieving of information written in a language different from the language of the user's query

a user types in English to retrieve documents written in French

CLIR systems use automatic translation techniques

Information Retrieval



Cross-language Information Retrieval



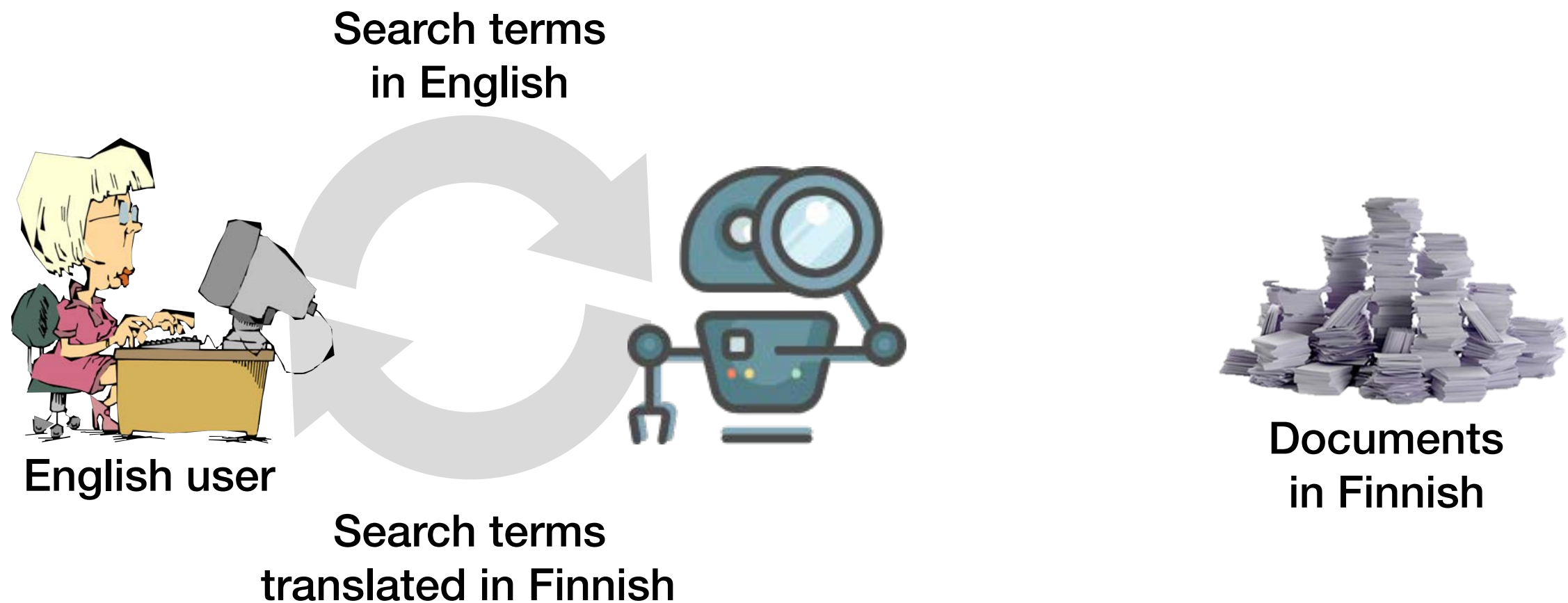
English user



Documents
in Finnish

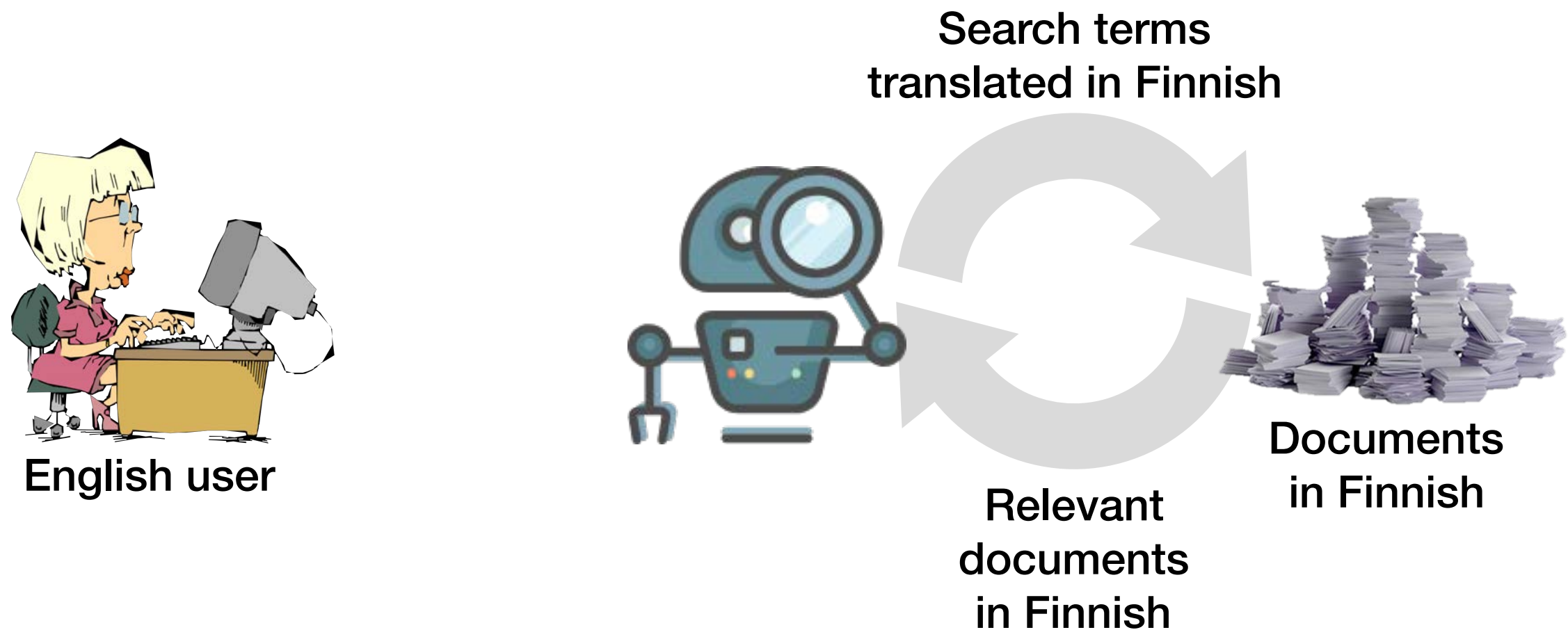
Cross-language Information Retrieval

step 1: query translation



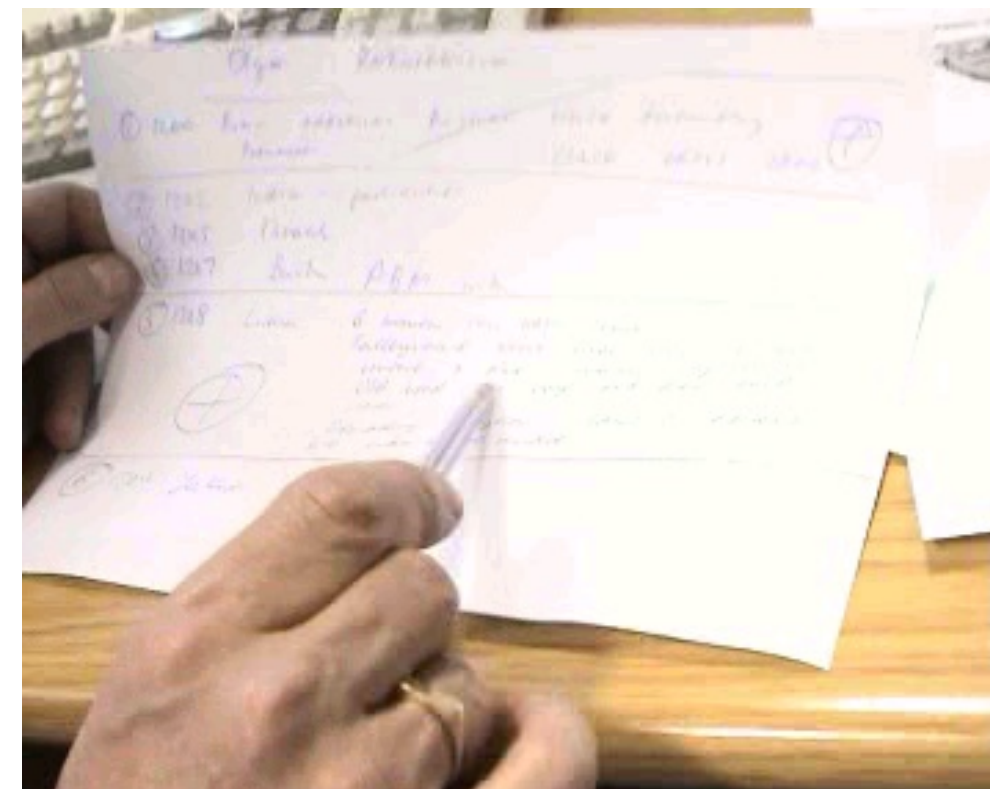
Cross-language Information Retrieval

step 2: search and retrieval

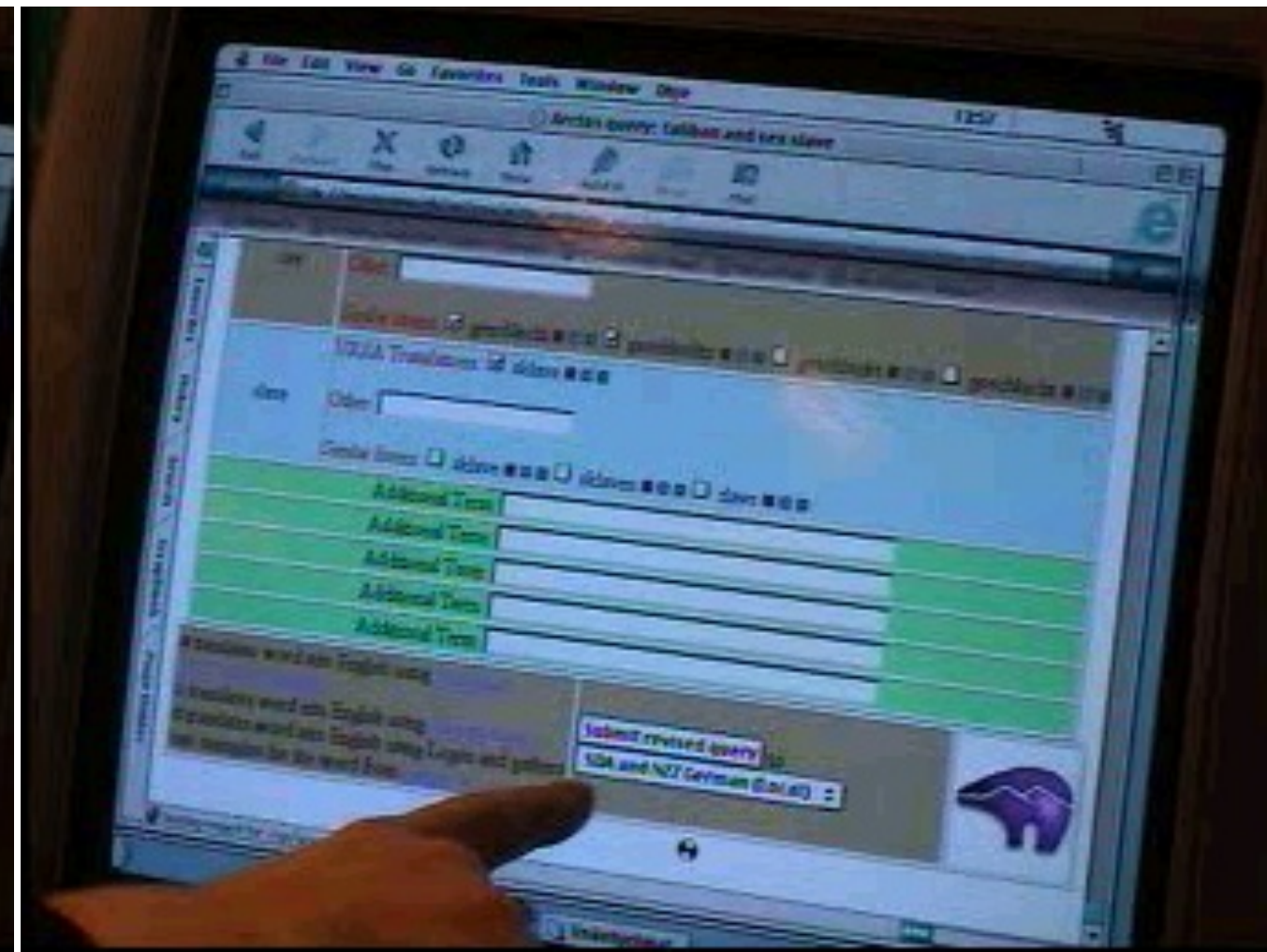
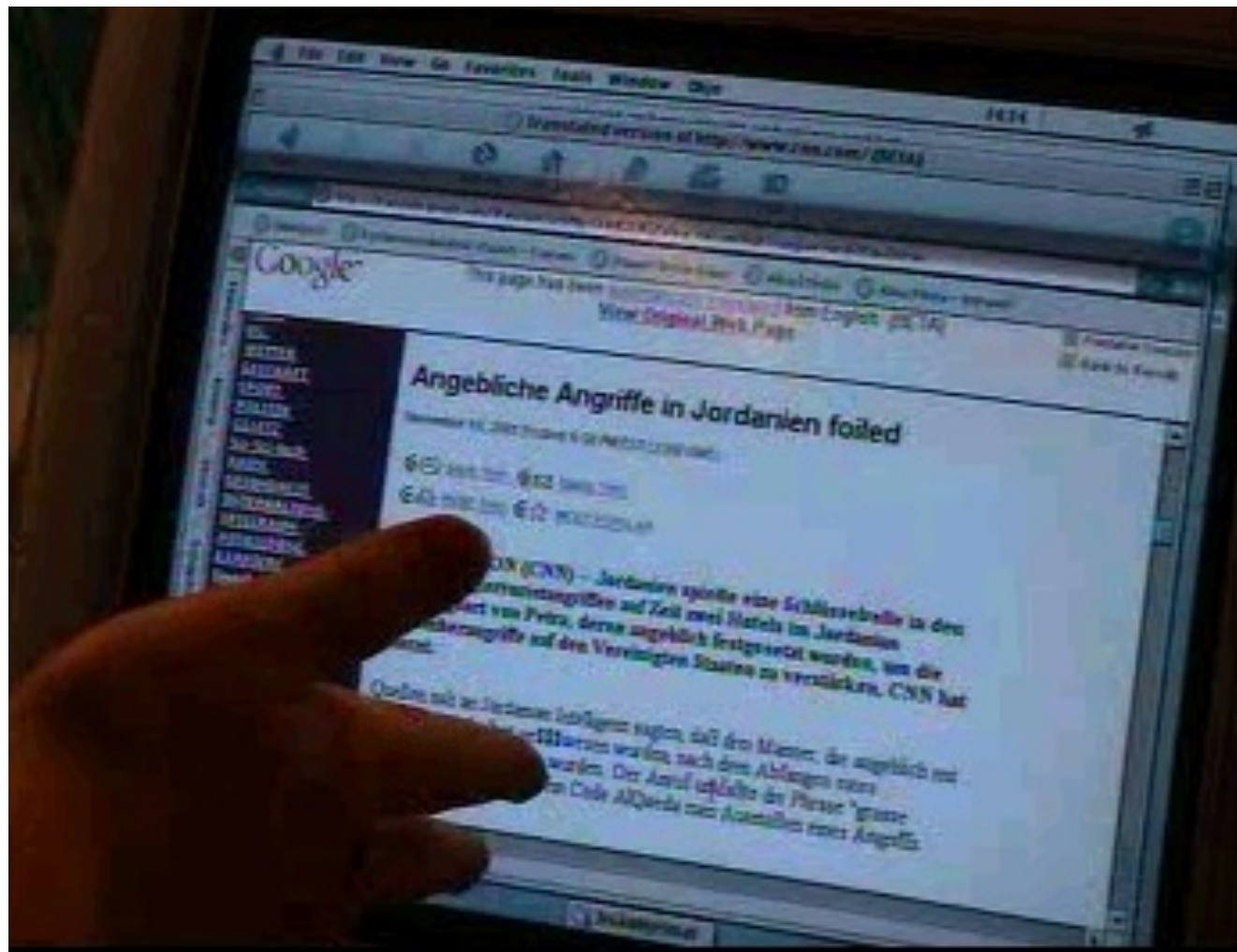


Who is the user?
What does s/he want to do?
Why?

Ethnographic study (2001): observations, interviews, questionnaires



Ethnographic study (2001): evaluations of existing tools



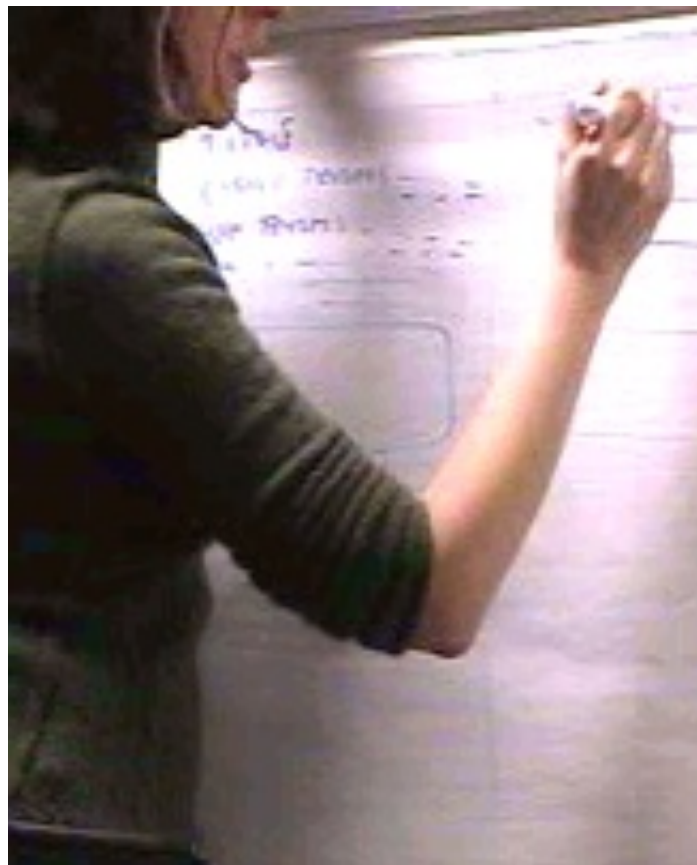
Ethnographic study (2001): participatory design of user interaction



SEARCH IN : ALMA-DB
FOR : TEXT (ANY) , IMAGES

G8 GENOA DISORDER

OTHER POSSIBLE WORDS		CURRENT QUERY TERMS
WORLD TRADE COMMERCE	→	G8 GENOVA
	←	ITALIA (DISORDER : CHAOS)
		SOTTOSSA (DISORDER : CHAOS)



RANK	LANGUAGE	CONCEPT	TYPE	DATE
• NO GLOBAL				
TITLE IN THE DOCUMENT LANGUAGE				
LANGUAGE : ENGLISH				
QUERY TERMS : G8, DISORDER - CHAOS				
KEY TERMS : FAIR TRADE, WORLD TRADE				
SUMMARIES - ENGLISH : - A THREE LINES -				
- SUMMARY IN ENGLISH -				
- A -				
- FINNISH : - A THREE LINES -				
- SUMMARY IN -				
- USER'S LANGUAGE -				
RELATED DOCUMENTS : TITLE ₅ IT IS A HYPERLINK				
TITLE ₁₀ IT IS A HYPERLINK				
• G8				
TITLE IN DOCUMENT LANGUAGE				
LANGUAGE : ITALIAN				
QUERY TERMS : G8				
KEY TERMS :				
SUMMARIES - ITALIAN : - A SUMMARY IN -				
- DOCUMENT -				
- LANGUAGE -				
- FINNISH : - A SUMMARY IN -				
- USER LANGUAGE -				

Journalists

what: receive brief / team work
search / check news

why: gain international perspective

how: browse reliable sources

needs: widen their language skills
ok with automatic translation

Business Analysts

what: extensive search across many
sources in many languages

why: prepare detailed reports

how: own knowledge/language skills

needs: access of multilingual sources
ok with automatic translation

Translators (BBC World)

what: monitor world-wide news

why: translate into English

how: listen to news channels
polyglots (up to 7 languages)

needs: retrieve news reports

Parliament Librarians (Finland)

what: retrieve laws / reports

why: provide a service to MPs

how: own/others language skills
search strategies (e.g.code)

needs: automatic translation - NO!

Comparative evaluation: translation check vs. translate-search

Netcape CLARITY - user testing

clarity search

Search in: ☒ English ☐ Finnish

For documents in: ☐ English ☒ Finnish

Date range: last two weeks Max documents: 10

racism bigotry immigration religion intolerance

☐ Translate ☒ Search

Translated terms:

racism:	rasismi: (racism)
racism:	rotusorto: (racial discrimination, racism)
bigotry:	fanatismi: (fanaticism, zealotry, bigotry)
bigotry:	kiihkeisyys: (fanaticism)
immigration:	maahanmuutto: (immigration)
immigration:	siirtolaisuus: migration, emigration, immigration
immigration:	perusturvakontrolli: (passport control)
religion:	uskonto: (religion)
religion:	uskonto: (religion, denomination)
religion:	oppi: doctrine, dogma, theory tenet, learning, erudition, scholarship, etc.
religion:	usko: belief, faith
intolerance:	intoleranssi: (no translation)
intolerance:	sietokyvyttömyys: (no translation)
intolerance:	suvaitsemattomuus: (intolerance, narrow-mindedness)
campaign:	kampanja: (campaign)
campaign:	tempaus: (wrench pull, snatch move, campaign, masterstroke)
campaign:	ekspedisi: (campaign, military expedition)
campaign:	operaatio: (operation, act of war)
campaign:	kampanjoita: (campaigns)

list report overview

sorted by rank, sort by language, sort by date

[Suosittu toissijainen](#)

05.05.2002 09:02, 132 words long, search terms: racism bigotry immigration religion intolerance campaign, terms found: racism

proper names: Thel Johansson Jari Nena Baroni Tilaus Matta Quark Deaally Marcel Tuuska Tai France Ammety-talonen TUOMI Toimittaja Al Ajaxin Tazzan-Juho Seminaari SYSTEM Kalle Laitanen Manta Ajax HINES RASMO Elina Rotunorum Peik Marhu Tazzan Innohuus Tila

keywords: viikko (week) laulu (song) erä (wilderness) rasti (race) ilta (evening) nälkä (hunger)

Netcape CLARITY - user testing

clarity search

Search in: ☒ English ☐ Finnish

For documents in: ☐ English ☒ Finnish

Date range: last two weeks Max documents: 10

racism bigotry immigration religion intolerance

☐ see query translated

list report overview

sorted by rank, sort by language, sort by date

[Suosittu toissijainen](#)

05.05.2002 09:02, 179 words long, search terms: racism bigotry immigration religion intolerance campaign, terms found: racism

proper names: Thel Johansson Jari Nena Baroni Tilaus Matta Quark Deaally Marcel Tuuska Tai France Ammety-talonen TUOMI Toimittaja Al Ajaxin Tazzan-Juho Seminaari SYSTEM Kalle Laitanen Manta Ajax HINES RASMO Elina Rotunorum Peik Marhu Tazzan Innohuus Tila

keywords: viikko (week) laulu (song) erä (wilderness) rasti (race) ilta (evening) nälkä (hunger)

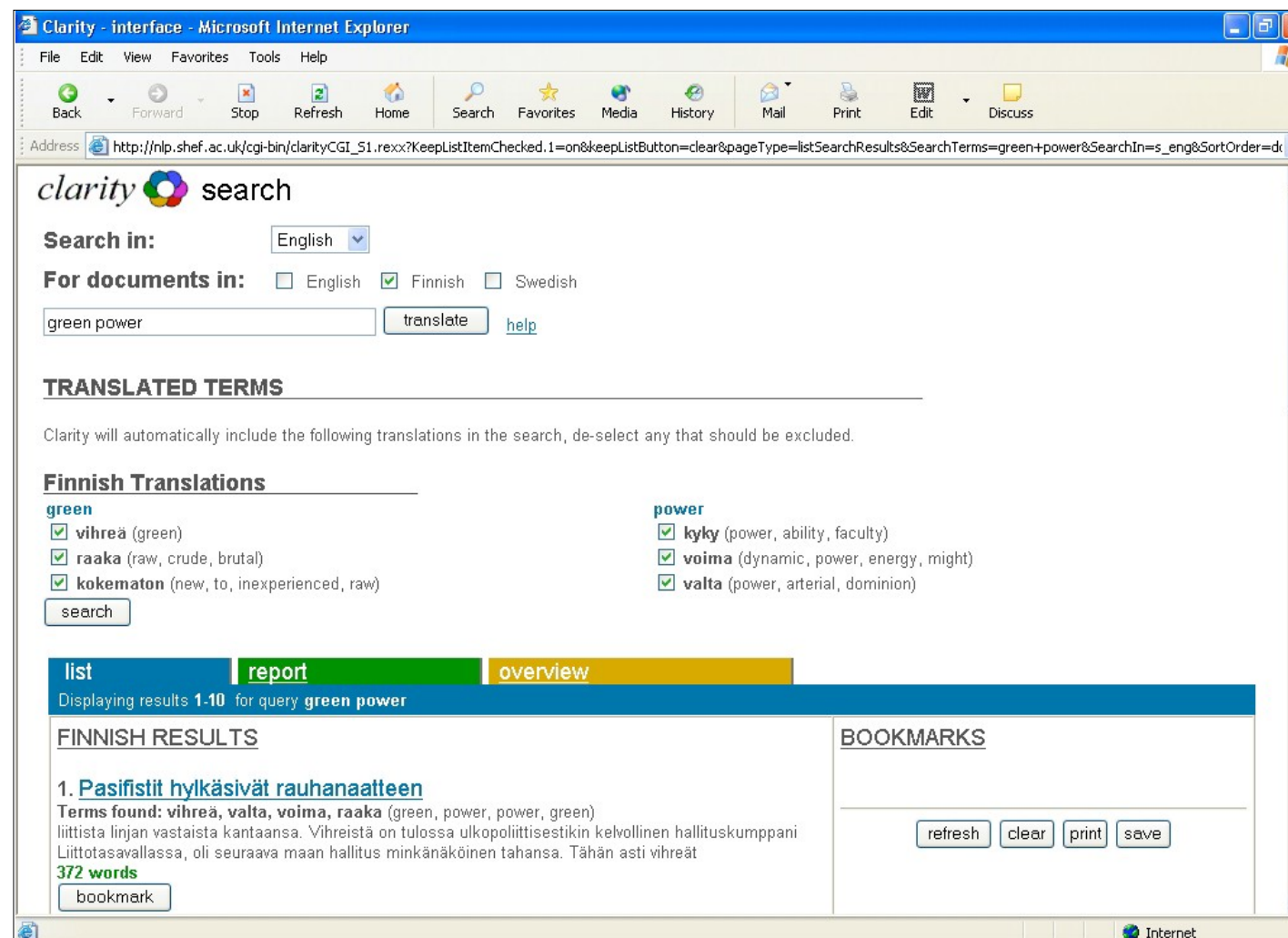
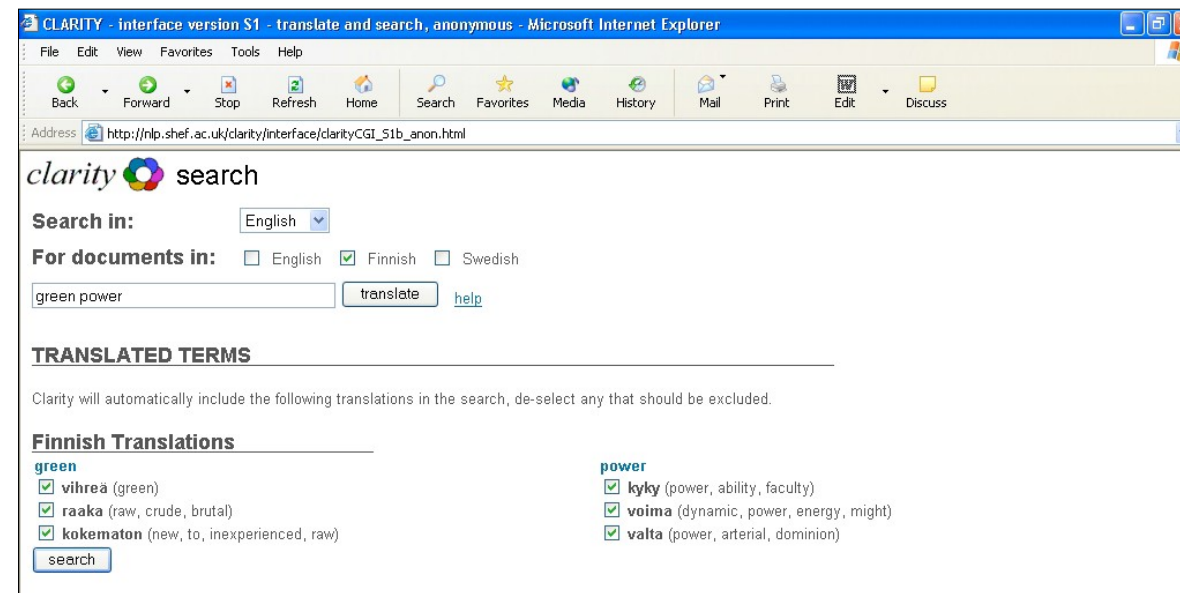
green power

green(s)

1. colour
2. grass
3. vegetables
4. unexperienced
5. party
6. golf course
7. not harmful to the environment
8. ... up to 13 different meanings

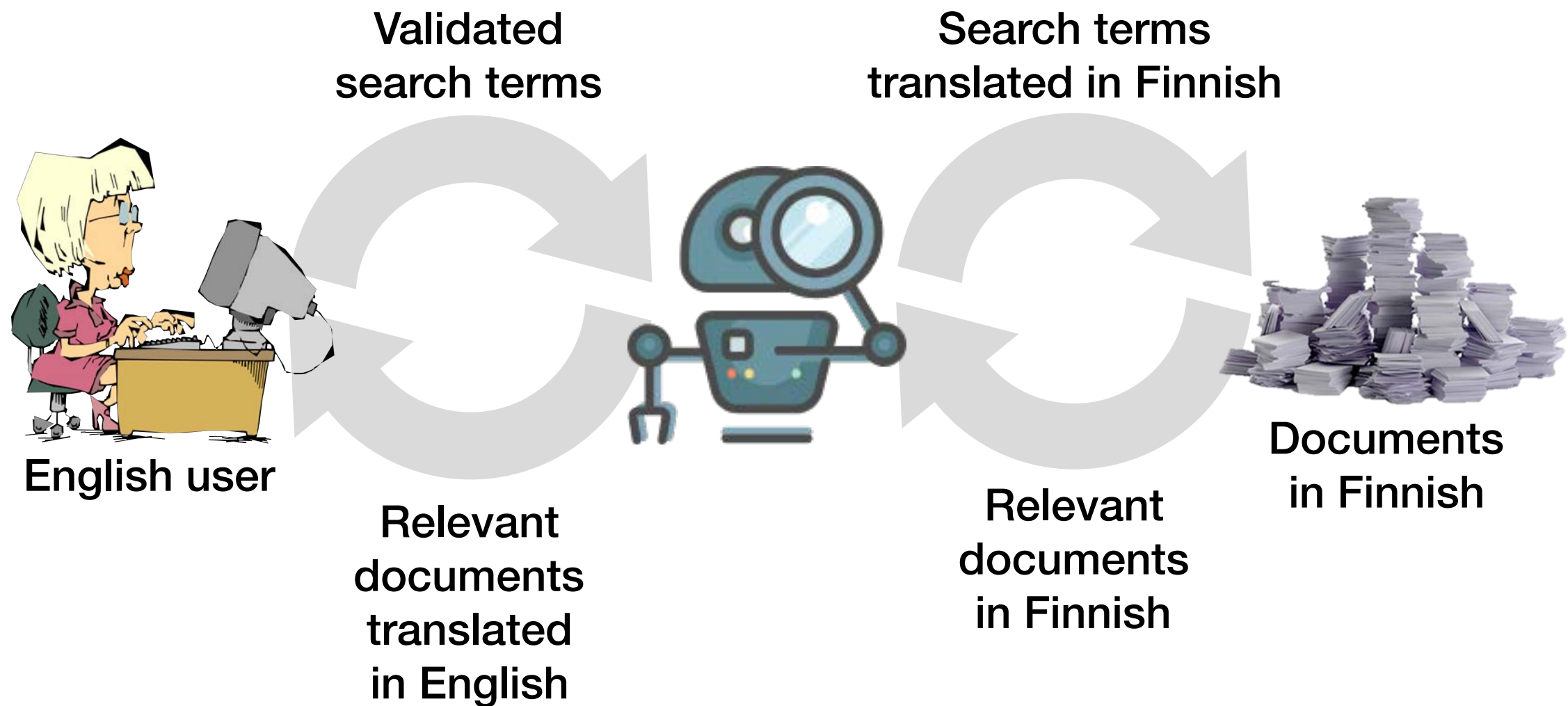
power

1. ability
2. authority
3. strength
4. influential
5. supernatural
6. performance
7. energy
8. ... + maths



Cross-language Information Retrieval

option 1: check translation, then search

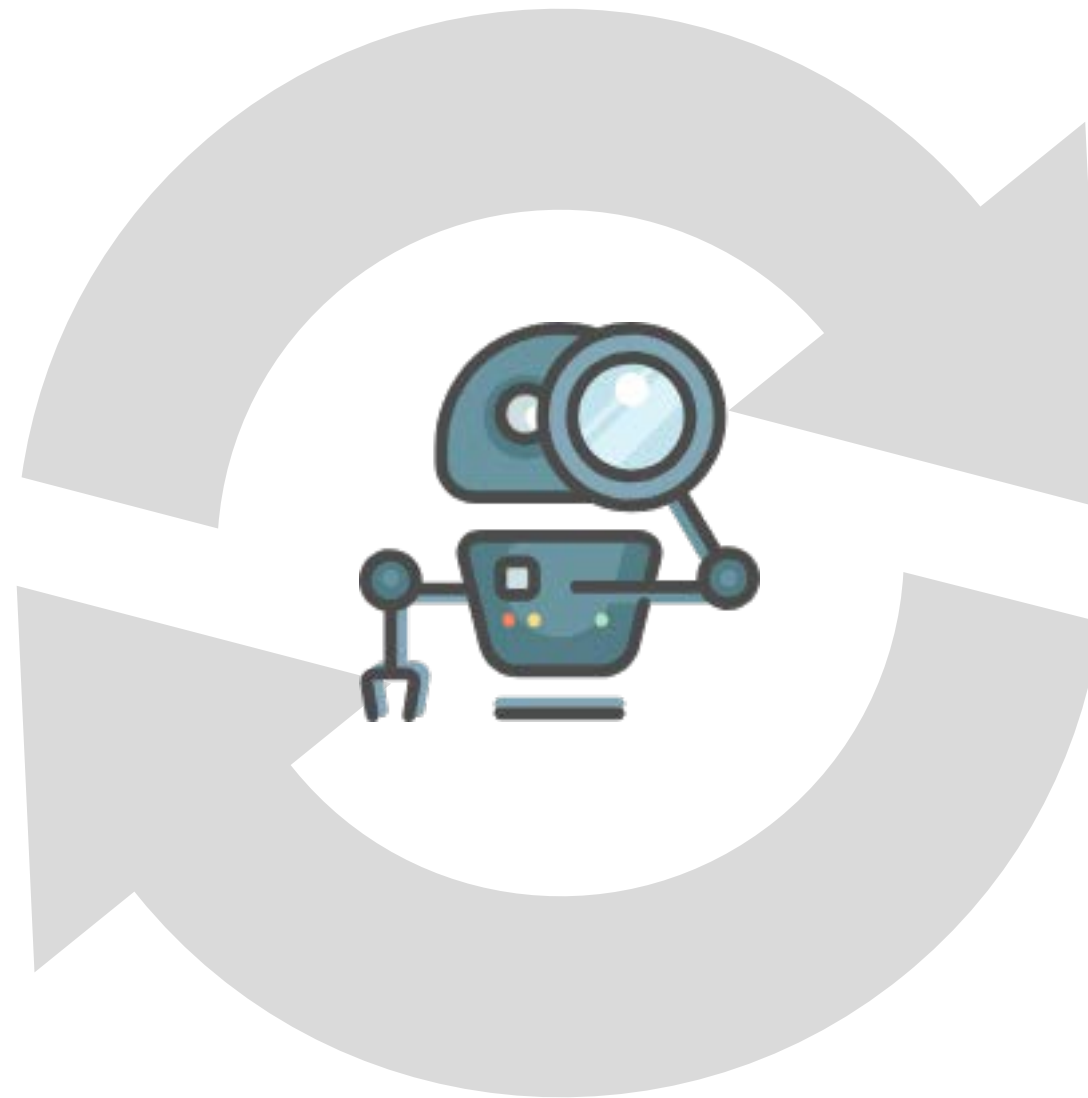


Cross-language Information Retrieval

option 2: translate & search

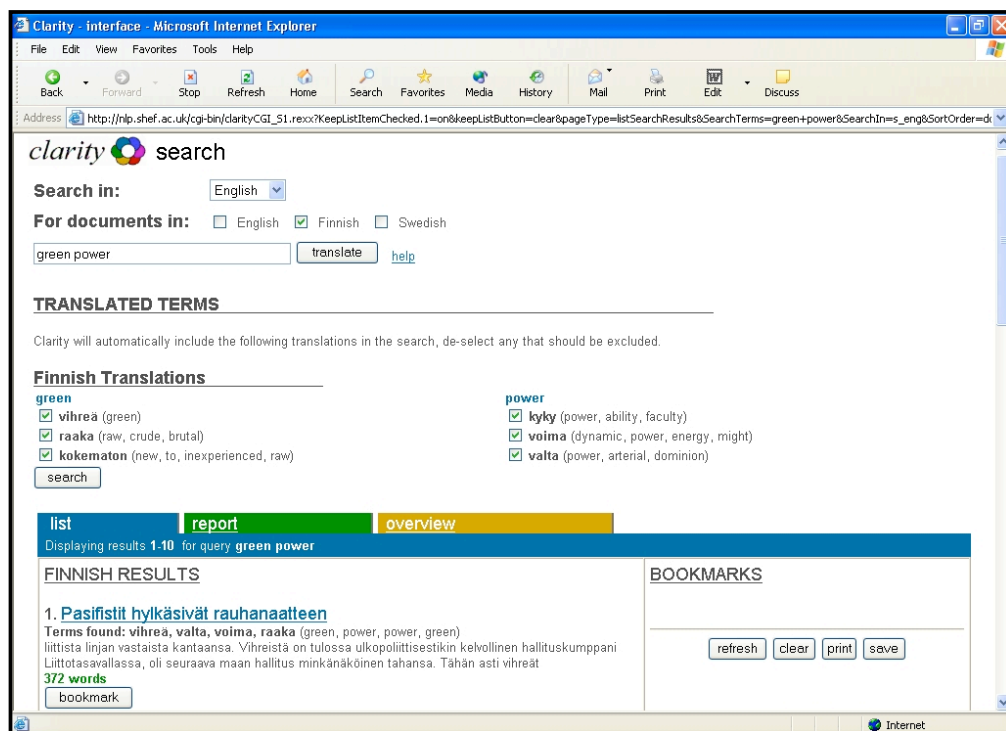
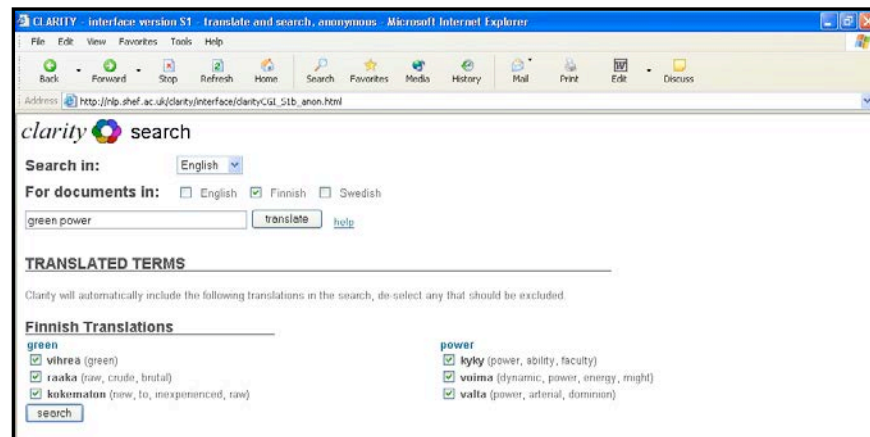


English user

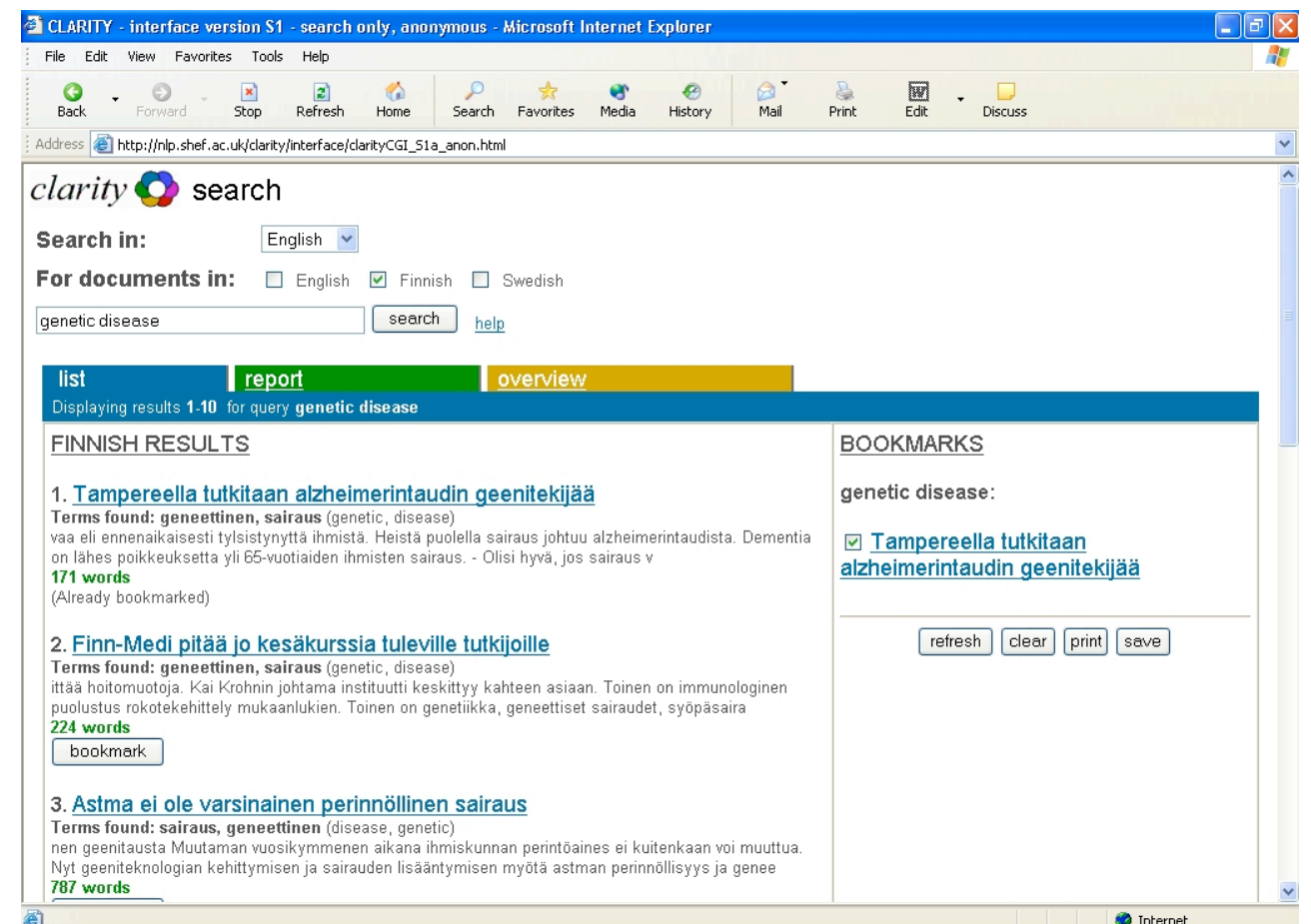


Documents
in Finnish

Comparative usability test: translation check vs translate&search

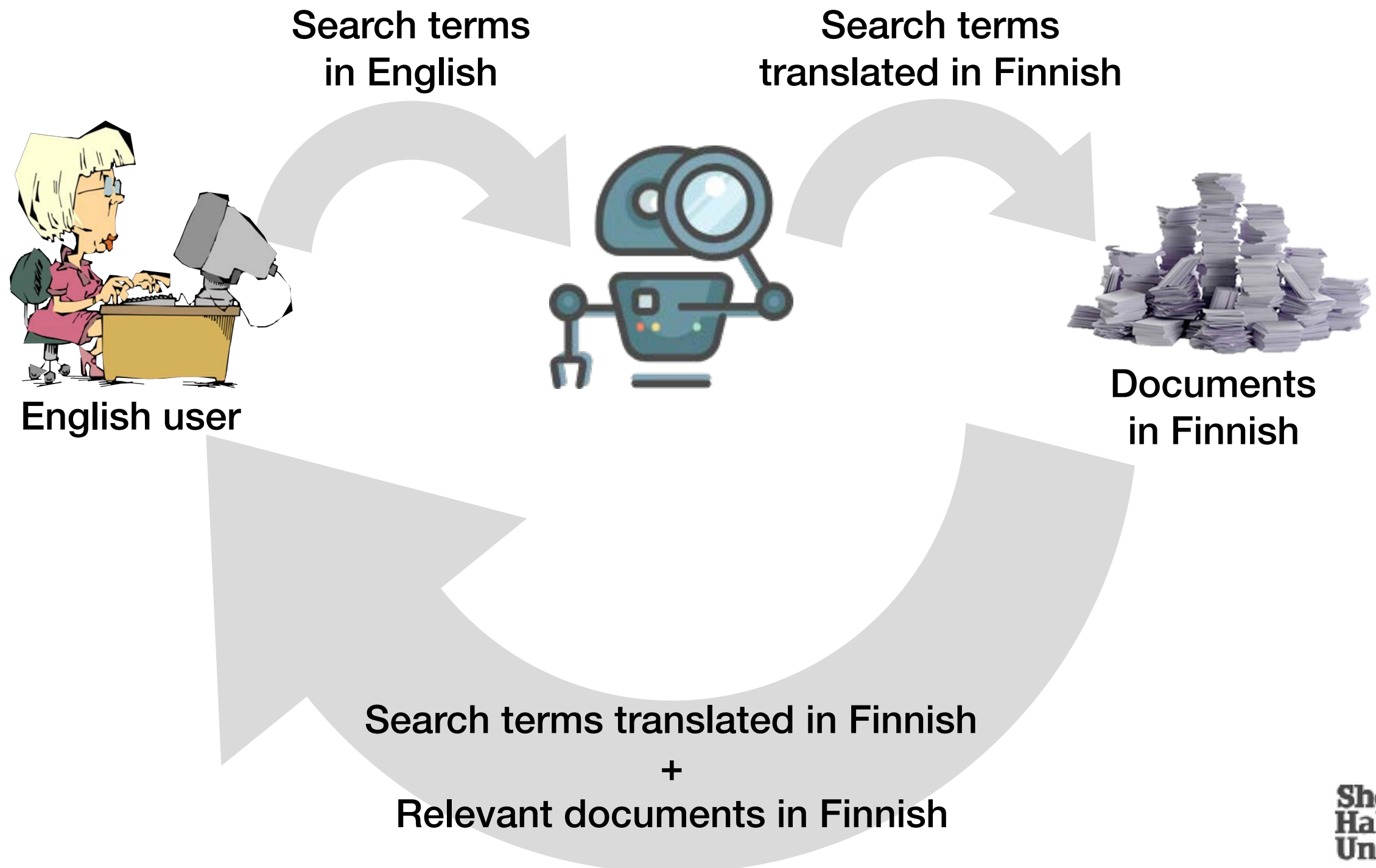


More effective and efficient



Preferred by users (user satisfaction)

Cross-language Information Retrieval



Clarity - interface - S1.9 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit Discuss Messenger

Google Address http://nlp.shef.ac.uk/cgi-bin/clarityCGI_S1.9.rexx?PageType=listSearchResults&ListStart=1&SearchIn=s_eng&docsInFin=on&searchTerms=prison+sentences+drug+dealing

clarity search

Search in: English

For documents in: ☐ English ☒ Finnish ☐ Swedish ☐ Latvian

prison sentences drug dealing [help](#)

Non-UK keyboard characters: ä, Å, Ä, å, ö, Ö, ð, Ø, ù, Û

TRANSLATED TERMS

Clarity automatically includes the following translations in the search, de-select any that should be excluded and click 'update' to refresh the results.

Finnish Translations

prison	sentences	drug	dealing
<input checked="" type="checkbox"/> vankila (prison, jail)	<input checked="" type="checkbox"/> virke (sentence)	<input checked="" type="checkbox"/> lääke (medicine, medicate, drug)	<input checked="" type="checkbox"/> käydä (attend, ferment, visit)
	<input checked="" type="checkbox"/> tuomio (judg, ment, conviction, verdict)	<input checked="" type="checkbox"/> huume (narcotic, drug, dope)	<input checked="" type="checkbox"/> kauppaa (shop, commercial, deal)
	<input checked="" type="checkbox"/> tuomita (condemn, denounce, adjudicate)	<input checked="" type="checkbox"/> huumata (dope, drug)	<input checked="" type="checkbox"/> kauppa (shop, commercial, deal)
			<input checked="" type="checkbox"/> jako (division, deal, partition)

list **summaries** **overview**

Displaying results 1-10 for query **prison sentences drug dealing**

FINNISH RESULTS

- Huumenuorelle kaksi vuotta vankeutta Hämeenlinnassa**
Huumenuorelle kaksi (two) vuotta (year) vankeutta (captivity) Hämeenlinnassa
Terms found: huume, tuomita, käydä, vankila, tuomio (drug, sentences, dealing, prison, sentences)
huumeidenkäytöstä ja kävi vankilassa häneltä sitä tivaamassa. Mies tunnusti ja sai lisätuomion. Nuorukainen on käyttänyt huumeita jatkuvasti kahden vuoden aikana, vankilassakin. Siellä hän velkaantui
58 words
- Ratsiamylly pitää huumeet loitolla Kylmäkoskelta**
Ratsiamylly pitää (give) huumeet (narcotic) loitolla Kylmäkoskelta
Terms found: huume, vankila, käydä, jako, kauppa, tuomio, tuomita (drug, prison, dealing, dealing, dealing, sentences, sentences)
keskusvankilaa, nuorisovankila, yhdeksän lääninvankilaa, Vilppulan ja Kestilän varavankilat sekä Sukevan kurssikeskus. Avolaitoksina toimii kolme varavankilaa, seitsemän avovankiosastoa ja kymmenen
703 words
- Ankarat tuomiot huumeekuriireille**
Ankarat (severe) tuomiot (judg) huumeekuriireille
Terms found: tuomio, huume, tuomita, kauppa, vankila (sentences, drug, sentences, dealing, prison)
tuomiot huumeekuriireille Yhdeksän kiloa heroiinia Thaimaasta Suomeen salakuljettanut pariskunta tuomittiin torstaina Vantaan käräjäoikeudessa vankeuteen törkeästä
390 words

BOOKMARKS

Internet





+



Bill Gates

*write * ...* to bypass the translation*

Dean Vietor
The New Yorker



Browser Di Frase

st (930)



uomini (40)



timpano (37)



chapel (33)



protezione (30)



- 12 documenti per i andrews della st del cattedrale e "il orientale di timpano"

☒ cattedrale ☒ st ☒ andrews ☒ "orientale di timpano"

[[controlla tutti](#) | [Uncheck tutto](#)]

St Andrews. Regola del regulus della st o della st. Chapel e Towe

St Andrews. Regola del regulus della st o della st. Chapel e Towe
 onnetta, con il timpano orientale della cattedrale.

timpano orientale (12)



corvo (8)



costruzione
dell'università (6)



headstone (5)



Segno: 15.874

ola del regulus della st o della st. Chapel e Towe

St Andrews. Regola del regulus della st o della st. Chapel e
 onnetta, con il timpano orientale della cattedrale.

Segno: 15.871

Rovine della cattedrale. st Andrews. Vista del timpano orientale



signora vestito bianco



St Andrew University - special collection

translated query: Mrs. dressed white man

Topic	Success (%)	Users	Average number of queries	Average number of terms
1	75	6	3.9	6.5
2	62.5	5	3.1	7.1
3	62.5	5	5.1	7.4
4	100	8	1.1	4.2
5	12.5	1	7.5	7.9
6	87.5	7	2.6	4.7
7	12.5	1	7.1	9.2
8	75	6	4	6.2
9	100	8	1.6	4.9
10	12.5	1	7	10.2
11	62.5	5	5.6	8.9
12	100	8	4.5	8.4
13	37.5	3	5.1	8
14	100	8	1	4
15	62.5	5	3.2	5.5
16	62.5	5	4.2	8.2



Add to Selection ☐

ID:

JV-R-5508

Title:

Setting Drag Nets, Lough Neagh.

Other Title:

Lough Neagh, Northern Ireland. Setting Drag Nets.

Place/Location:

[World, Europe, Northern Ireland](#)

Creation Date:

1958

Created By:

[J Valentine & Sons](#)

Available Format + Media:

photographic prints / gelatin silver prints

Subject Index:

[fishing equipment \(145\)](#)

[fishermen \(144\)](#)

[rowing \(749\)](#)


[harbours \(962\)](#)

[rowing boats \(741\)](#)

Cataloguing Details

Description:

Clinker built rowing boat, with two fishermen, one casting nets and the other rowing; low harbour wall with slipway.

Add to Selection 

ID:

Cataloguing Details



Add to Selection ☐

ID:

JBM-33-2-25

Title:

Montu's Temple, Nag-el-Medamoud.

Other Title:

Nag-el-Medamoud. Temple of Montu.

Place/Location:

[World, Africa, Egypt](#)

Creation Date:

1992

Created By:

[Milne, Mr, James, Burt, \(1918 - 1997\)](#)

Available Format + Media:

positives (photographs) / slides (photographs)

Subject Index:

[buildings - stone \(8220\)](#)

[temples \(305\)](#)

[Egyptian ruins \(208\)](#)

Cataloguing Details

Designing the linguistic interaction

1. users do not use well formed queries and machine translation fails (but bilingual dictionaries fail in dealing with bi-grams , e.g. adjective + noun)
2. users pick up terms displayed by the system in follow-up queries, both correct translations and ill-translated terms
3. all metadata should be indexed as they can be the key to succesful retrieval, e.g. location
4. users search for what is prominent in the image but describing the background improved the retrieval for a subset of images
5. archival metadata is key and extra fields such as foregraound and background descriptions could improve performance

UMBERTO
ECO



Mouse or Rat?

TRANSLATION AS NEGOTIATION

Daniela Petrelli
Art & Design Research Centre
Sheffield Hallam University

d.petrelli@shu.ac.uk

@DanielaPetrell1