

Improving interaction with the user in Cross-language Question Answering through Relevant Domains and Syntactic Semantic Patterns

iCLEF-2004



Universitat d'Alacant
Universidad de Alicante



Departamento de
Lenguajes y Sistemas Informáticos



Contents

1. Objective.
2. Experiment.
3. Method I: Relevant Domains.
4. Method II: Syntactic Semantic Patterns.
5. Results.
6. Conclusions.



Objective

- How to assist users in the localization of the correct answer in passages written in a language different from the one of the query.
- Two interaction methods:
 - Relevant Domains
 - Syntactic semantic patterns

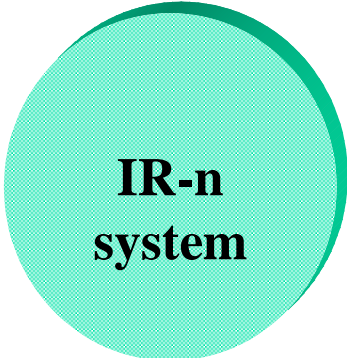


Experiment

1. Query formulation and translation to English.
2. Extraction of relevant passages with IR-n System.
3. Interaction with the users and localization of the answer.

Method I: Relevant Domains

Method II: Syntactic Semantic Patterns



**IR-n
system**



**Rel.
Domains**



SSP

Information Retrieval systems

General Custer was Civil War Union Major soldier. One of the most famous and controversial figures in United States Military history. Graduated last in his West Point Class (June 1861). Spent first part of the Civil War as a courier and staff officer. Promoted from Captain to Brigadier General of Volunteers just prior to the Battle of Gettysburg, and was given command of the Michigan "Wolverines" Cavalry brigade.

He helped defeat General Stuart's attempt to make a cavalry strike behind Union lines on the 3rd Day of the Battle (July 3, 1863), thus markedly contributing to the Army of the Potomac's victory (a large monument to his Brigade now stands in the East Cavalry Field in Gettysburg). Participated in nearly every cavalry action in Virginia from that point until the end of the war, always performing boldly, most often brilliantly, and always seeking publicity for himself and his actions. Ended the war as a Major General of Volunteers and a Brevet Major General in the Regular Army.

Upon Army reorganization in 1886, he was appointed Lieutenant Colonel of the soon to be renown 7th United States Cavalry. Fought in the various actions against the Western Indians, often with a singular brutality (exemplified by his wiping out of a Cheyenne village on the Washita in November 1868).

His exploits on the Plains were romanticized by Eastern United States newspapermen, and he was elevated to legendary status in his time. The death of his friend, Lucarelli change his life.

IR systems study in a global way to determine the similarity between a query and a document



Query

Passage retrieval systems

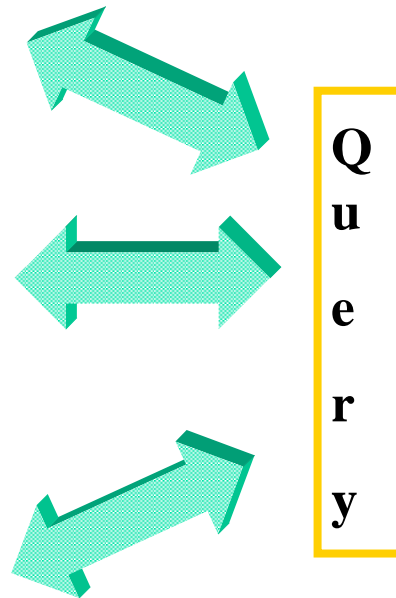
General Custer was Civil War Union Major soldier. One of the most famous and controversial figures in United States Military history. Graduated last in his West Point Class (June 1861). Spent first part of the Civil War as a courier and staff officer. Promoted from Captain to Brigadier General of Volunteers just prior to the Battle of Gettysburg, and was given command of the Michigan "Wolverines" Cavalry brigade.

He helped defeat General Stuart's attempt to make a cavalry strike behind Union lines on the 3rd Day of the Battle (July 3, 1863), thus markedly contributing to the Army of the Potomac's victory (a large monument to his Brigade now stands in the East Cavalry Field in Gettysburg). Participated in nearly every cavalry action in Virginia from that point until the end of the war, always performing boldly, most often brilliantly, and always seeking publicity for himself and his actions. Ended the war as a Major General of Volunteers and a Brevet Major General in the Regular Army.

Upon Army reorganization in 1886, he was appointed Lieutenant Colonel of the soon to be renown 7th United States Cavalry. Fought in the various actions against the Western Indians, often with a singular brutality (exemplified by his wiping out of a Cheyenne village on the Washita in November 1868).

His exploits on the Plains were romanticized by Eastern United States newspapermen, and he was elevated to legendary status in his time. The death of his friend, Lucarelli change his life.

PR systems divide the document into passages and evaluate the relevance of each passage



We can obtain the most relevant part of a document

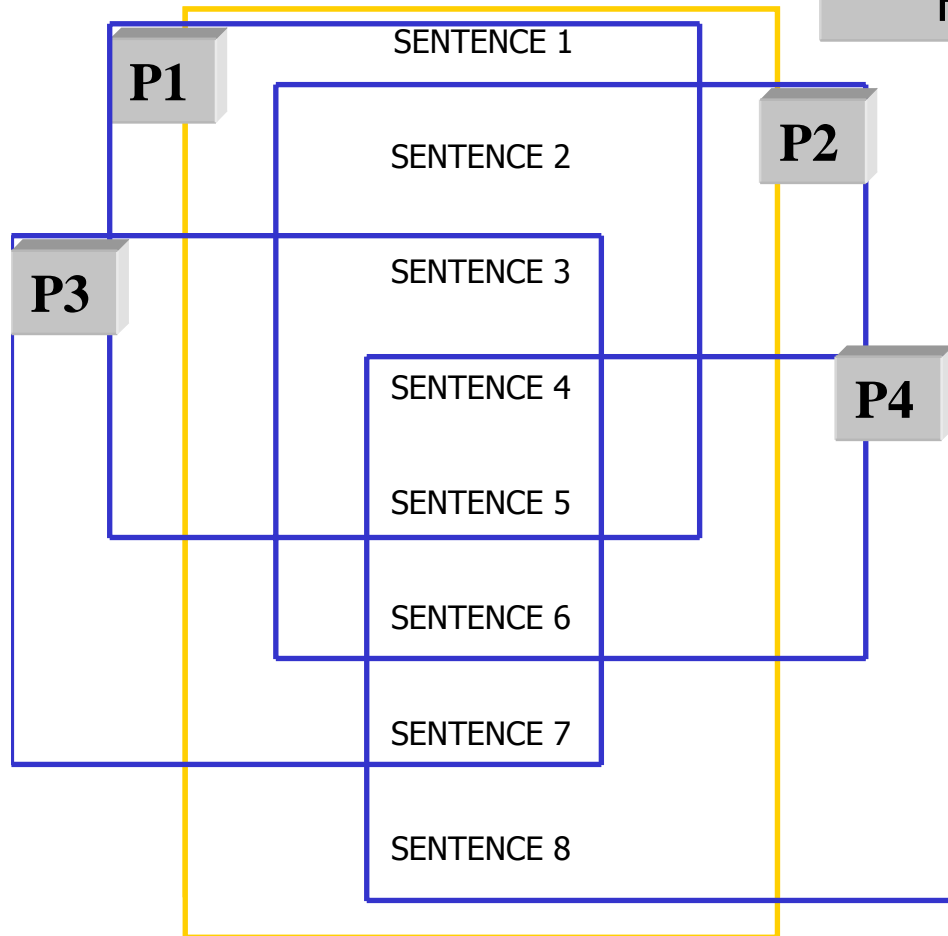


IR-n system

- IR-n system use the sentence to define the passages
 - Every passage have the same number of sentences
- Sentences are full units allowing to show an understandable information by users or provide this information to a subsequent system (like QA system)

IR-n system (2)

How works the IR-n system

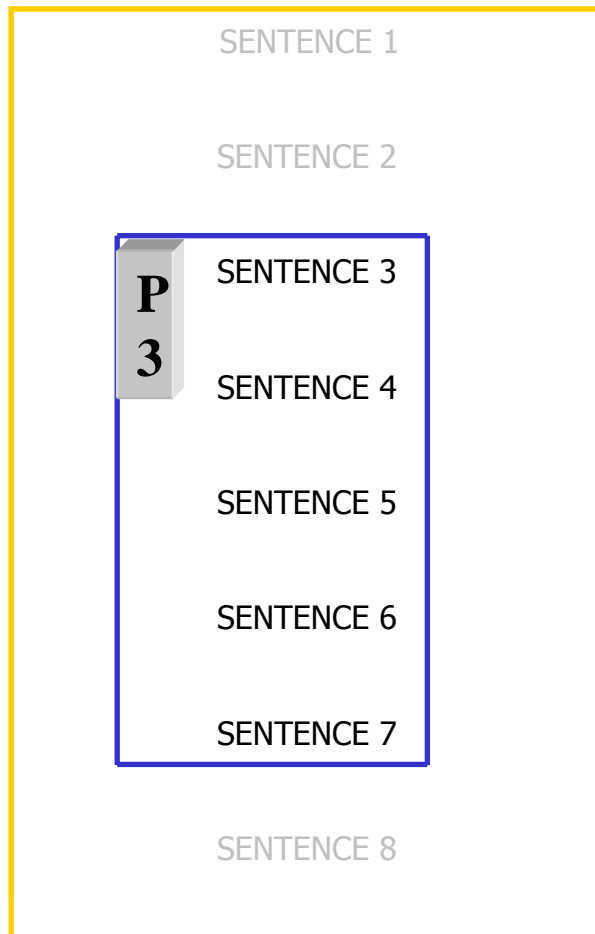


1 – It obtains the sentences of document

2 – It defines the passages using the size of passages and degree of overlapping

IR-n system (3)

How works the IR-n system



1 – It obtains the sentences of document

2 – It defines the passages using the size of passages and degree of overlapping

3 – It obtains the more relevant passage of each document



IR-n system (4)

- Previous experiences with IR-n system in QA task
 - TREC QA task 2001 and 2002
 - Clef QA task 2003 and 2004
- IR-n system selects the most relevant passages for a QA system. In this way the QA system only process short passages (and not full documents).
- In iCLEF task we use only the more relevant passages for each query.

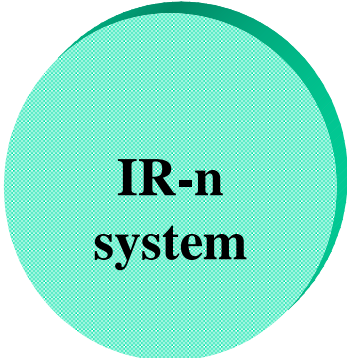


Experiment

1. Query formulation and translation to English.
2. Extraction of relevant passages with IR-n System.
3. Interaction with the users and localization of the answer.

Method I: Relevant Domains

Method II: syntactic semantic patterns



**IR-n
system**



**Rel.
Domains**



SSP



Relevant Domains

- The more relevant and representative ontological domains of a word.
- Extracted automatically from WordNet Domains.
- Example:
 - **Question:** Who is the managing director of FIAT?
 - **Relevant domains:** Administration Economy Transport
- The passages order has been recalculated according to the similarity between the relevant domains of the query and the relevant domains of the passage.



Relevant Domains

- The system shows user:
 - Query in Spanish and its relevant domains in English.
 - Passage in English and its relevant domains.
- Theoretically, the passage with the correct answer must contain the same or similar relevant domains than the query.
- The user can decide previously if the passage contains the answer or not.

Query in Spanish

Query Domains

Passage in English

Passage Domains

Pregunta 17: (1 de 2)

¿Quién es el director gerente de FIAT?

[pregunta: administration economy transport]

DOMINIOS:

- doctrines
- telecommunication
- physics
- transport
- economy
- pedagogy
- publishing
- sexuality
- commerce
- sociology

The Synergie is essentially the same base vehicle as its Peugeot and Fiat rivals.

But each company can instal its own engines and settle on its own equipment levels, with the result that these MPVs are by no means identical when you get them out on the road.

Citroen manages to ring the changes in the Synergie by offering two different engines, a 125bhp two-litre petrol job and a 92bhp 1.9-litre turbo diesel, as well as five, six or seven-seater styles, and equipment specifications at LX, SX and VSX levels.

There are nine models on the list, at prices from #16,200 to #18,755 for the petrol versions, and from #17,200 to #23,090 for the turbo diesels.

Unexpectedly, the top ranked VSX is diesel rather than petrol.

Solución:

[NO SE LA SOLUCIÓN! |Siguiente pregunta! >

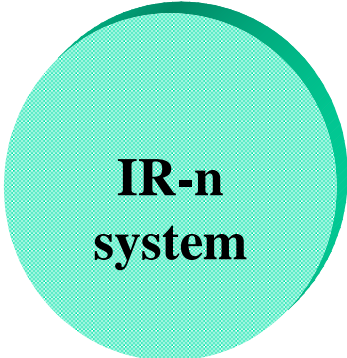


Experiment

1. Query formulation and translation to English.
2. Extraction of relevant passages with IR-n System.
3. Interaction with the users and localization of the answer.

Method I: Relevant Domains

Method II: Syntactic Semantic patterns



**IR-n
system**



**Rel.
Domains**



SSP



Syntactic Semantic Patterns

- Formed by:
 - the verb of a sentence and its sense,
 - the subcategorization frame: the head of the main complements and their senses.
- Extracted automatically from the passages.
- Example:
 - **Sentence:** "However, the finance ministers and central bankers from the so-called Grup of Seven who heard the Russian' presentation voiced skepticism that the ambitious goals can be met."
 - **SSP:**
 - voice (finance banker group skepticism)
 - hear (Russian presentation)
 - meet (goal)



Syntactic Semantic Patterns

- Hypothesis:
 - Only the most important information of each sentence is shown to the user:
 - the most important words of each sentence –the verb and the subcategorized nouns– and the syntactic and semantic relation between them.
 - Users with low competence in English can decide if the answer is in the passage or not.

Query in Spanish

Pregunta 19: (3 de 20)

¿Quién es el ministro ruso de Finanzas?

Pasaje 1:

SS Patterns

PATRONES:

voice (finance banker group scepticism)
 hear (russians presentation)
 meet (goal)
 give (day g-7 projection)
 follow (europe united recession world)
 talk (meeting u.s. japanese hour)
 fail (headroom)
 be (focus g-7 reform election election)

Passage in English

However, the finance ministers and central bankers from the so-called Group of Seven who heard the Russians' presentation voiced skepticism that the ambitious goals can be met. .
 Earlier in the day, the G-7 ministers gave a cautiously optimistic projection that Europe may soon follow the United States out of one of the worst recessions since World War II. .
 Also, at a separate, bilateral meeting, U.S. Treasury Secretary Lloyd Bentsen and Japanese Finance Minister Hirohisa Fujii talked for about an hour but failed to make any headway in resolving the issues that have led to a major trade dispute between their two nations. .
 But the main focus of Saturday's G-7 meeting was on the Russian reform process and the obstacles that threaten it in the wake of last December's parliamentary elections -- elections that saw proponents of a free-market economy trounced by ultranationalists and former Communists. .

Solución:

Siguiente

¡NO SÉ LA SOLUCIÓN! ¡Siguiente pregunta! >

Pregunta 9: (19 de 21)

¿Cuándo murió Lenin?

When Lenin dies?

Pasaje 9:

PALABRAS:

die (lenin 1924)

begin (work evening)

go (berezhkov bed morning)

continue (dawn)

mean ()

unnerve (stalin berezhkov)

say (berezhkov)

appear (dictator)

fill (berezhkov desire)

hide (child)

be (berezhkov stalin)

be (dictator axiom man)

be (war day)

arrive (washington)

tell (berezhkov)

recall (berezhkov)

Lenin died in 1924, began work late in the evening and often continued until dawn, which meant that Berezhkov never went to bed until morning. . * . Stalin unnerved him, Berezhkov said, filling him with a desire to hide like a child behind the curtains when the dictator appeared.

Berezhkov was well aware of Stalin's fear of people who knew too much.

The dictator's favorite axiom was "A dead man cannot bite." .

When a war day was two days late arriving in Washington,

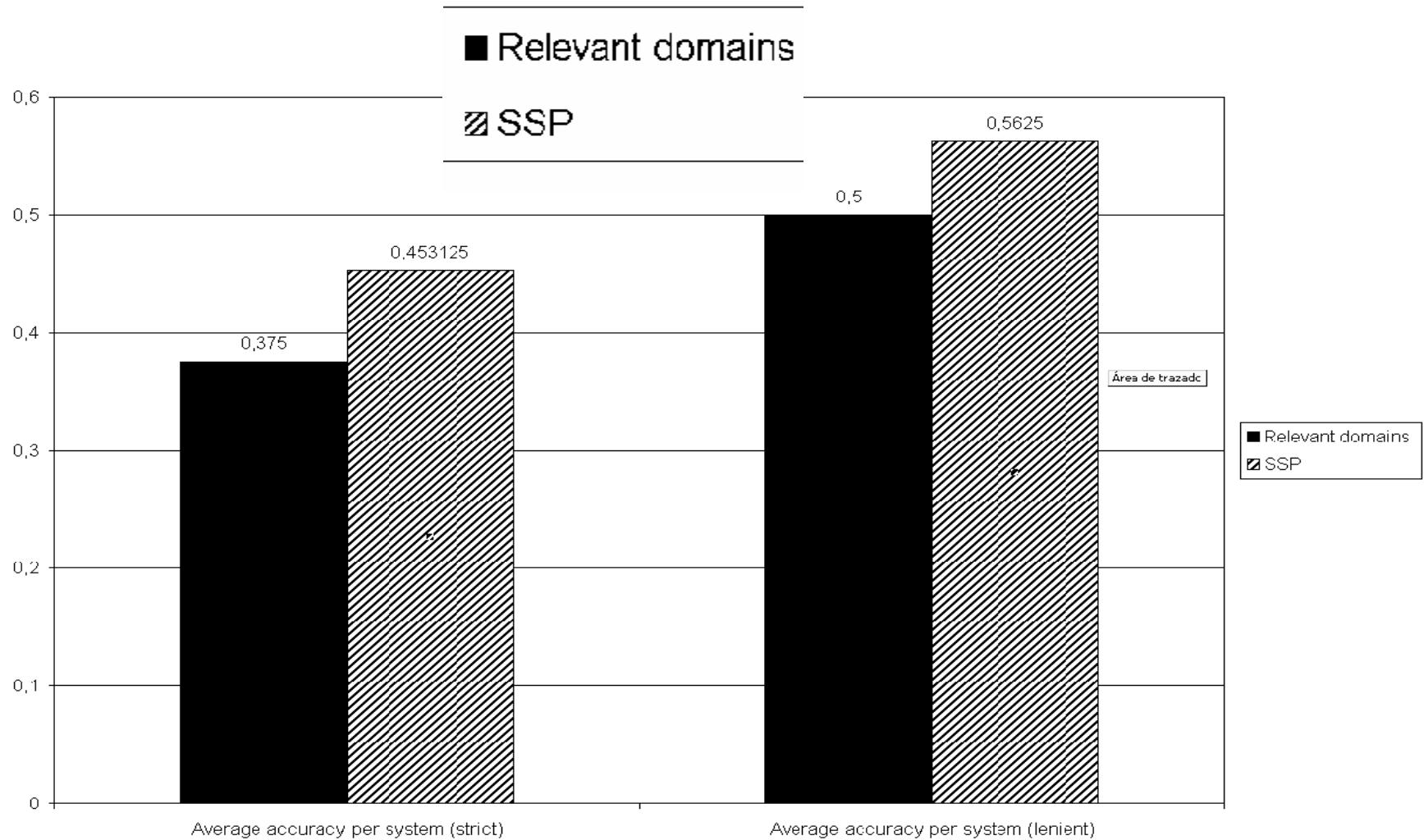
Berezhkov was unable to investigate but could turn up no mistakes, he recalls. .

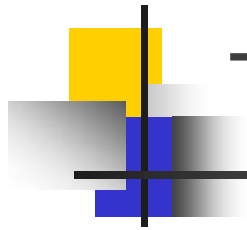
Die (Lenin 1924)

Solución

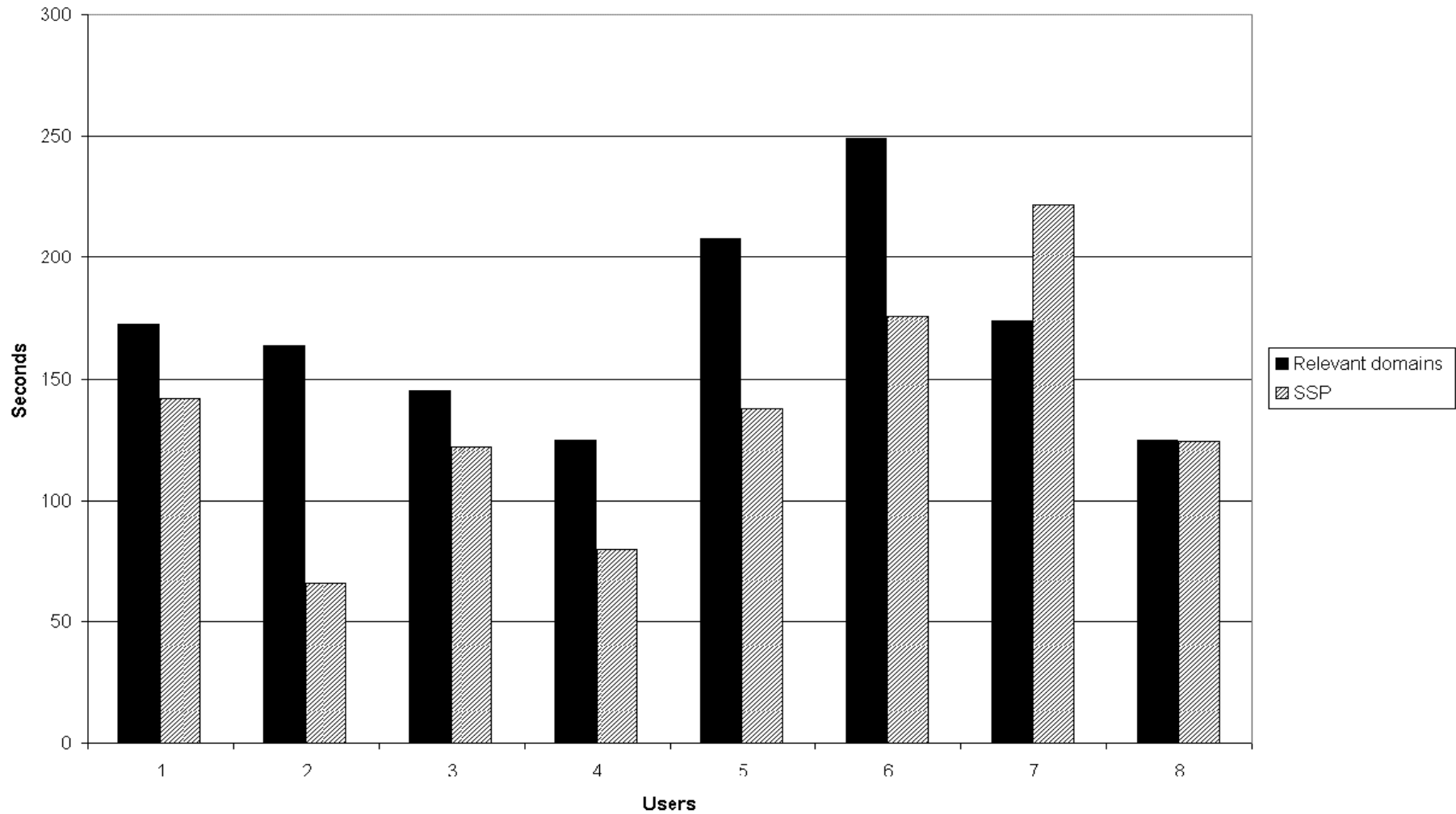
Siguiente

General results





Time consuming





Conclusions

- The order in which the passage are shown to the user in the SSP method (the output order of the **IR-n system**) seems to be more correct.
 - The order of passage based on the similarity of relevant domains do not seem the most correct one.
- According to the results and the personal opinion of the users, **relevant domains** method really help the localization of the answer.



Conclusions II

- The SSP method achieve better results, but the users said that it is very difficult to read the patterns
 - only nouns and verb, without any linguistic connection.
- It is necessary to improve the extraction of SSP in order to ensure that the patterns will contain the possible answers: if the answer do not appear in a SSP, the user do not locate it.



Work in progress

- We think we have to use some kind of translation.
 - In this sense, it is necessary some kind of translation (at least, superficial translation) to really help the localization of the answer



**Improving interaction with the user in
Cross-language Question Answering
through Relevant Domains and
Syntactic Semantic Patterns**



iCLEF-2004