

# QA better than IR ?

**Dominique Laurent**  
Synapse Développement  
33 rue Maynard  
Toulouse, France  
dlaurent@synapse-  
fr.com

**Patrick Séguéla**  
Synapse Développement  
33 rue Maynard  
Toulouse, France  
patrick.seguela@syn  
apse-fr.com

**Sophie Nègre**  
Synapse Développement  
33 rue Maynard  
Toulouse, France  
sophie.negre@synaps  
e-fr.com

## Abstract

A Question Answering (QA) system allows the user to ask questions in natural language and to obtain one or several answers. If compared with a classical IR engine like Google, what kind of key benefits QA bring to users and how to measure their distinctive performances. This is what we shall attempt here to determine, specially in providing a comparative weak and strong points table of each system, along with showing how QA systems, in particular our Qristal QA system, requires up two to six time less “user effort”.

## 1 Introduction

Asking questions in natural language and obtain short answers (if possible the exact answer) make QA systems the paramount of Information Retrieval.

Through TREC and CLEF international campaigns, Question Answering systems are evaluated both in monolingual and multilingual use (e.g. Voorhees, 2005; Vallin, 2005). Nevertheless, very few comparisons of performances took place between Question Answering systems and Information Retrieval engines (Kwok, 2001; Radev, 2002; Buccholz, 2002; McGowan, 2005). If so, they mainly focus on the quality of the supplied answers and on the time for user to obtain the answer.

We hereby attempt to define an evaluation method to compare performances and user-friendliness of both Question Answering systems and Information Retrieval engines. We will apply this method to our system Qristal and to the Google Desktop Search<sup>1</sup> engine..

## 2 Background

### 2.1 Qristal pedigree

Qristal (French acronym of "Questions-Réponses Intégrant un Système de Traitement Automatique des Langues", which can be translated by "Question Answering System using NLP") is, as far as we know, the first Multilingual Question Answering System available on the consumer market (B2C). It handles French, English, Italian, Portuguese or Polish.

Qristal allows the user to query on a static corpus or on the Web. It supplies answers in one or any of the 4 languages.

Our system is described in detail in other papers (Amaral, 2004; Laurent 2004; Laurent 2005-1; Laurent 2005-2). Qristal is based on our Cordial syntactic analyzer and extensively uses all the usual constituents of the natural language processing, while, as seldom found, remarkably featuring anaphora resolution and metaphor detection.

Originally developed within the framework of the European project TRUST<sup>2</sup> and M-CAST<sup>3</sup>, our system has evolved, over the last five years, from a monolingual single-user program into a multilingual multi-user system.

### 2.2 Qristal benchmarks and performances

A beta version of Qristal was evaluated in July 2004 in the EQueR<sup>4</sup> evaluation campaign organized in France by several ministries (Ayache, 2004; Ayache, 2005). With a MRR ("Mean Reciprocal Rank"; cf. Ayache, 2004) of 0.58 for the exact answers and 0.70 for the snippets, our system ranked first out of the seven

<sup>1</sup> Google Desktop Search : <http://desktop.google.com/>

<sup>2</sup> IST-1999-56416 <http://www.trustsemantics.com>

<sup>3</sup> EDC 22249, <http://www.m-cast.infovide.pl>

<sup>4</sup> <http://www.technolanguen.net/article61.html>

Question Answering systems evaluated.

The marketed version of Qristal was evaluated during the CLEF 2005 (Laurent, 2005-2) and obtained 64% of exact answers for French to French, 39.5% from English to French and 36.5% from Portuguese to French. Once again, Qristal ranked first in this evaluation, for French engines and for all cross language systems, all pairs considered.

Since this evaluation, the resources were increased and some algorithms revised, so our last tests brought us a 70% of exacted answers and 45 % for cross language.

### 3 QA and IR

It is true that, intrinsically, IR engines and QA systems differ in design, objectives and processes. An IR engine is geared to deliver snippets or docs from a query, a QA system strive to deliver the exact answer to a question.

If one is to differentiate 3 key features of both systems, one of the first difference concerns the query mode : natural language for the QA systems and “Boolean like” for the IR engines. We define “Boolean like” extensively as the use of Boolean operators associated to underlying constraints induced by the word matching techniques. The table 1 gives the results of Google Desktop for natural language requests and Boolean requests (set of questions detailed below) and we can see that results with natural language requests are not so good:

rank of answers	Google snippets natural language requests	Google snippets Boolean requests	Google snippets or docs natural language requests	Google snippets or docs Boolean requests
1	4.2 %	9.7 %	7.3 %	30.6 %
1-5	6.7 %	18.5 %	14.2 %	52.1 %

Table 1. Results with Google Desktop Search on natural language and Boolean requests

This performance table shows that classical engines are not suited to answer questions in natural language. To quote Google “A Google search is an easy, honest and objective way to find high-quality *websites* with information relevant to your search.”

The Google technology considers, at least in French, equally and of same “weight, words like, “de” or “le” and the highly semantically-loaded words of the query. This leads to a dramatic upsurge of noise in their results. Therefore, using

classic engines require a good knowledge of their syntax and their underlying word matching techniques, like the necessity of grouping between quotation marks, the “noun phrases” and the expressions.

The second difference concerns what is delivered to the user. Question Answering systems deliver one or more exact answers to a question and their context whereas classical engines return snippets with links to the texts those snippets were extracted from.

The third difference relates to the dynamic and openness status of the corpora. Usually QA systems use confined or close corpora with low up-date rate, while classical IR engines are tuned to the Web queries and their reference file are continuously updated.

Qristal QA is able to deliver answers from both web-based queries and closed corpora. We were eager to apply our proposed metrics on the web-based deliveries, but unfortunately, we had not at our disposal the appropriate web reference file of questions and answers, probably impossible to elaborate considering the extremely high up-dating rate of the web pages.

Therefore we had but no choice to use a closed corpus, Google Desktop being able to manage this type of corpus (see note 1). We used the reference file of questions and answers established for EQueR.

#### 3.1 Question Answering systems

Qristal interface mimics almost all the usual screen template of IR engine (see figure 2). It displays results in different languages, keeps the track or trace of the precedent requests, allows the user to choose the requested corpus and, if thought necessary, to make a semantic disambiguation of the question terms.

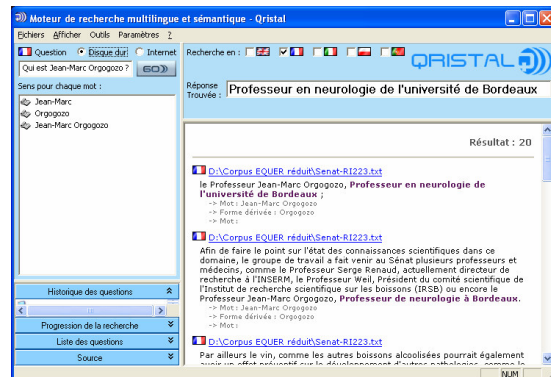


Figure 2. QRISTAL

Results (answer, sentences, links) are displayed at three levels: if an exact answer is found, it

is displayed in the top right part of the window, sentences in the lower right part of the window and links on top of the sentences. Note that the words or phrases supporting the inferred answer are put in bold in the text. These words are sometimes pronouns (anaphora) and, frequently, synonyms or derivate forms of the request words.

### 3.2 Information Retrieval engines

Numerous Information Retrieval engines are available for closed or web-based corpora.

For our evaluation we selected the Google engine, as it is available both for a web and closed PC desktop usage (in version "Desktop Search"), although, regrettably, this beta version had a few minor defects. The snippets supplied by Google are generally fragments of sentences, sometimes with cut words, stemming from excerpts of text seeming to correspond the best to the query. All you can expect is a help in the selection of the text(s) in which is likely to be present the answer to your query rather than a pinpointed and elaborated exact answer.

### 3.3 Evaluation method of the performances

#### Corpus of requests and answers

The corpus selected for this evaluation is the corpus used for the EQueR evaluation campaign. This choice was justified by the size of the corpus (over half million texts for about 1.5 Gb) and, especially, by the fact that we have an important corpus of questions (500) with many answers and references for all these questions.

To generate a comprehensive package of tests of Question Answering systems, ELDA, organizer of the EQueR campaign, made a compilation of all the results returned by all the participants. Then, thanks to a thorough examination by several specialists (one of whom being an author of this paper), this corpus has been verified, increased and validated. In doing so, most certainly, the immense majority of the possible answers are inventoried, the great majority of the references of these answers is known<sup>5</sup> to the extent that this corpus of questions and answers can be automatically run, the results subsequently requiring only a reduced amount of checking.

The initial corpus of 500 questions has been reduced to 330 questions for this evaluation. In

fact, the last 100 questions were only reformulations of previous questions and offered not enough interest. 30 questions concerned binary answers YES/NO and 40 questions concerned lists as answers. As Information Retrieval engines are not able to return binary or list answers, including them within the evaluation would have biased it. Finally, five questions were without any answer. They were removed by the organizers of EQueR and we also did so accordingly. For these five "no-answer" questions, Qristal systematically returned correct answers i.e. NIL, where as a classical search engine like Google would systematically return at least one answer. We decided not to include these "NIL" questions so as not further penalize the IR Google engine.

#### Evaluation of the "user effort"

We have two competing systems on the same corpora and a reference file with questions and answers. We need now to define the basis of their comparison.

The main comparative evaluation between the Question Answering systems and the Information Retrieval engines (Kwok, 2001) considered only the reading time while counting characters to be read to reach the answer. Knowing the delay needed to obtain the results in most Question Answering systems (McGowan, 2005), it seems necessary to take also in account this delay if we want to measure the global user effort to obtain an answer to his question.

We consider that the user wants a correct answer to his question and we consider that the answer is correct if this answer can be found in the snippet or in the text linked to the snippet for Google (to the sentence for Qristal). So we compared the quality of the systems as follows: percentage of correct answers ranked first, ranked within the five first, the ten first and the hundred first. We considered both the answer as part of the snippets or part or the snippets and documents.

For the user, the quality of the answers returned, especially in the first results page, is paramount. But, we think another item has to be taken into account: the time needed to obtain the answer. This time is the compound of three elements:

- the time to key in the question,
- the delay before the results display,
- the reading time of the snippets or sentences to reach a correct answer.

---

<sup>5</sup> a set comprising textual corpora, questions and answers is available at ELDA (<http://www.elda.org/article139.html>)

Addition of these three elements provides the measure of the "user effort".

### Time to key in the question

This time is shorter for a "Boolean like" engine. Typing a question in Qristal needs in average nine seconds more than with the query in Google. However it supposes that the user types the Boolean like request at the same speed than a natural language request. This implies that the user is very familiar with the Google syntax. For example the question 6:

*Quel âge a l'abbé Pierre ?*

will be converted into Google syntax by:

*"abbé Pierre" ans*

This Boolean request increases the probability to obtain the effective age, not only snippets with the words "âge" and "abbé Pierre". Other example, the question 37:

*Quel évènement a eu lieu le 27 décembre 1978 en Algérie ?*

will be converted into Google syntax:

*"27 décembre 1978" Algérie*

This Boolean request is needed either words like "évènement" or "avoir lieu" will bring out more noise than correct answers.

So, we translated the complete list of question in the Google syntax in addition to the questions set in natural language. Compared results were shown on Table 1.

To measure the time necessary to enter questions, we counted the number of characters typed (always inferior in Google) and multiplied this number by an average speed of 150 characters by minute. We know that a professional typist types at a speed of 300 to 400 characters per minute, so our chosen speed corresponds to a user keying in with two fingers. The following table gives the numbers of characters and the times for the two systems:

	characters	mean time
Qristal	49.1	19.6 seconds
Google Desktop	27.2	10.9 seconds

Table 3: mean time of question entering

### Delay to display the results

This is the elapsed time between the click on the button "OK" and the display of the results. Note here that, strangely, Google Desktop has a response time distinctly bigger than the response time of Google on the Web, especially when the request contains a group of words between quotes.

### Reading time to reach one answer

To fix a reading speed, we tested several users. An average speed of 40 characters by second (2 400 characters per minute, or also 400 words per minute) seems a fair measure. It corresponds to a reader with a higher-education background, according to Richaudeau, 1977. While making these tests, we noted that, if the user knows the answer, the reading speed of both the snippets and texts would increase to 100 characters per second (6 000 characters per minute, or 1 000 words per minute). Even if very few questions have an obvious answer, we decided to calculate the reading times with both speeds of 40 characters per second and 100 characters per second.

The speed of 100 characters per second will however be considered as a superior limit that favours clearly Google where the snippets, constituted by fragments of sentences, sometimes fragments of words, are more difficult and longer to read than the sentences returned by Qristal.

## 4 Results of the benchmark

### 4.1 Evaluation on the 330 questions

For the 330 questions of the evaluation, results are:

Answer rank	Google Desktop snippets	Qristal snippets	Google Desktop snippets or docs	Qristal snippets or docs
Exact answer		69.7 %		69.7 %
1 or exact	9.7 %	81.8 %	30.6 %	86.7 %
1-5	18.5 %	88.2 %	51.9 %	94.5 %
1-10	21.5 %	88.8 %	58.5 %	96.1 %
1-100	27.0 %	90.9 %	70.0 %	98.5 %
Not found	73.0 %	9.1 %	30.0 %	1.5 %

Table 4: Percentage of correct answers / 330

Qristal returns an exact answer for nearly 70% of questions and a correct answer is returned as exact answer or in the first sentence for 82% of the 330 questions. This has to be compared with the 10% of answer found in first position by Google Desktop. If we consider the snippets and the documents, Qristal returns a correct answer in first rank for 86% of the questions and Google in more than 30%.

These results give a clear advantage to the Question Answering system on the Information Retrieval engine. This superiority in quality exists also in quantity. Here is the table of user efforts to obtain a correct answer:

	Google Desktop	Qristal
Type the question	10.9 s	19.6 s
Display results	3.0 s	2.1 s
Reading 40 char/second	59.3 s	7.1 s
Reading 100 char/second	23.7 s	2.8 s
Display results + Reading 40 characters/second	62.3 s	9.2 s
Display results + Reading 100 characters/second	26.7 s	4.9 s
Type the question + Display results + Reading 40 c./second	73.2 s	28.8 s
Type the question + Display results + Reading 100 c./second	37.6 s	24.5 s

Table 5: Mean times compared for the two systems (on 330 questions)

It appears that the time to type down the question in Qristal is nearly 9 seconds longer than with Google. The elapsed time before display is similar. On the other hand, as Google gives a correct answer in a higher rank or in a document, not a snippet, the number of characters to be read before reaching an answer is finally more important. Finally, if we consider the average reading speed of 40 characters per second, Qristal needs in average 29 seconds against 73 seconds, in other words the user effort to obtain a good answer is 2.5 times higher with Google than with Qristal.

If we don't take into account the time to enter the question, Google requires a user effort 6 to 7 time higher than Qristal to reach an answer. This comparison would be effective in the case of voice-based submitted query. In that case, the acquisition of the question would become more difficult according to the syntax of the Boolean engine ("open the quotes", "close the quotes"...) )

#### 4.2 Evaluation on 231 questions

Looking carefully at each answer returned by Google Desktop, we discovered that it ignored some texts or, more exactly, some parts of texts, especially the end of these texts. The help pages of this software point out this "bug" :

*However, if you're searching for a word within the file, please note that Google Desktop searches only about the first 10,000 words. In a few cases, Google Desktop may index slightly fewer words to save space in your search index and on your hard drive.<sup>6</sup>*

Of course this default impacted on the results and the comparisons. Thus, we decided, in a second iteration of this evaluation, to consider only the 231 questions where Google Desktop found at

least one correct answer. Google Desktop found no answer with those 99 (330-231) removed questions for two main reasons. Firstly, as it doesn't manage a full indexation of documents. Secondly, as some complex questions like "why" or "how" questions often lead it to silence on this evaluation.

This selection of 231 questions favours Google Desktop but it allows a more accurate comparison. Here are the results for those 231 questions:

Scale	Google Desktop bribes	Qristal bribes	Google Desktop bribes ou docs	Qristal bribes ou docs
Exact answer		73.6 %		73.6 %
1	13.9 %	89.6 %	43.7 %	90.9 %
1-5	26.4 %	94.8 %	74.0 %	97.4 %
1-10	30.7 %	94.8 %	83.5 %	97.8 %
1-100	38.5 %	96.1 %	100.0 %	99.6 %
Not found		3.9 %		0.4 %

Table 6: Number of answers and percentages / 231

The corpus of 231 questions is thus "easier" than that of 325 questions. This confirms the score of Qristal for the exact answers: 73.6% versus 69.7%, and the score for the correct answer in first rank: 89.6% versus 81.8%. But the results of Google are of course better: 13.9% in the first snippet against 9.7%, 43.7% in the first snippet or the first document, against 30.6%.

However the advantage of the QA system over the IR engine is clear in terms of quality, especially if we consider only the snippets. This advantage is also clear for the user effort, even if any of the answer not found by Qristal penalizes this system as the reading time of this question is the consolidation of all reading times of all the snippets displayed for this question!

	Google	Qristal
Type the question	10.1 s	18.7 s
Display results	3.5 s	2.0 s
Reading 40 char/second	28.4 s	3.3 s
Reading 100 char/second	11.4 s	1.3 s
Display results + Reading 40 char/second	31.9 s	5.3 s
Display results + Reading 100 char/second	14.9 s	3.3 s
Type + Display results + Reading 40 char/second	42.0 s	23.9 s
Type + Display results + Reading 100 char/second	24.9 s	21.9 s

Table 7: Mean times compared for the two systems (on 231 questions)

<sup>6</sup> <http://desktop.google.com/support/bin/answer.py?answer=24755&topic=209>

The mean times of question entering and displaying results are nearly the same for those 231 questions than for the 325 questions. But, because Google Desktop finds an answer to all questions, the reading time before a correct answer is, in that case, reduced for Google.

The mean times of question entering and displaying results are nearly the same for those 231 questions than for the 325 questions. But, because Google Desktop finds an answer to all questions, the reading time before a correct answer is, in that case, reduced for Google.

Finally, with an average reading speed of 40 characters by second, the user effort is two times higher with Google Desktop than it is with Qristal. And if we don't take into account the time to type in the question, the user effort with Google is 6 times higher than with Qristal.

Using the same presentation than Kwok, 2001, the following graph gives the compared results of the two systems. On Y-axis is the number of correct answers and in X-axis the number of characters read, for the 231 questions:

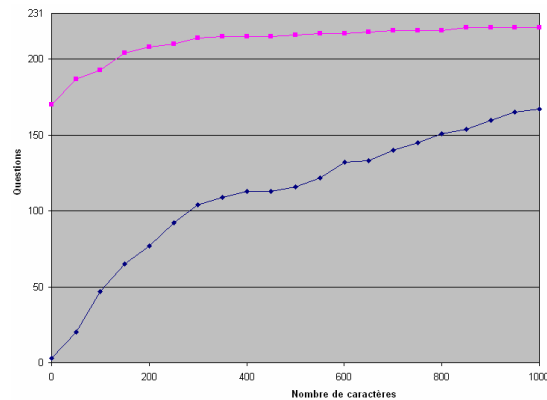


Figure 8 : number of correct answers by characters

The interest of Question Answering systems is particularly noticeable at the beginning of the graph, seeing that Qristal displays a correct answer as exact answer at the top of the screen in more than 70% of the questions while Google Desktop needs to read about 1000 characters in the snippets and in the documents to obtain a similar success rate.

### 4.3 Comparison by type of question(s)

The above statistics concern all types of queries. In fact, 25 questions wait a definition, the others being factual requests. The following table of the 231 questions corpus gives the results for these two categories:

Answer rank	Google snippets	Qristal snippets	Google snippets or docs	Qristal snippets or docs
Exact definition		64.0 %		64.0 %
Definition in rank 1	32.0 %	88.0 %	48.0 %	92.0 %
Definition in rank 1-5	52.0 %	96.0 %	68.0 %	100 %
Exact factual		74.8 %		74.8 %
Factual rank 1	11.7 %	89.8 %	43.2 %	90.8 %
Factual rank 1-5	23.3 %	94.7 %	74.8 %	97.6 %

Table 9 : Percentages of answers by type of questions

The only significant gap in these results is that Google provides better results for definitions (32% of correct answers in the first snippet against 12% for the factual questions).

We also looked at the questions beginning by "comment" ("how"), but we excluded those beginning by "comment s'appelle" ("how is called") or "comment est mort" ("how did somebody die"), i.e. 16 questions (3, 50, 59, 90, 93, 117, 148, 154, 165, 196, 199, 234, 247, 249, 263, 295). The results are :

Rank	Google snippets	Qristal snippets	Google snippets or docs	Qristal snippets or docs
Exact		25.0 %		25.0 %
1	0.0 %	56.3 %	18.8 %	56.3 %
1-5	6.3 %	68.8 %	25.0 %	81.3 %
1-10	6.3 %	68.8 %	25.0 %	81.3 %
1-100	12.5 %	68.8 %	37.5 %	81.3 %
Not found	87.5 %	31.2 %	62.5 %	18.7 %

Table 10: Percentage of correct answers for questions beginning by "comment" ("how")

These results are not satisfying for any of the two systems but Qristal displays a correct answer in 9 cases on 16, versus 0 for Google Desktop. This underlines that the Question Answering systems are more successful when the queries are not purely factual requests. Most certainly this could be caused by the fact that those questions require a deeper analysis.

A closer examination of factual questions revealed that the most difficult questions for the Information Retrieval engines are the questions about location. For example Google Desktop finds the country related to the Vilvorde town only at the 23rd rank, and the country related to Johannesburg is given only at the 13th rank; the region of Cancale is displayed at the 18th rank;

and the department (county) where is located Annemasse only at the 23rd rank.

More generally, a search engine finds answers more easily when these answers contain the words of the query. For example, to the query 252 ("*À quelle peine fut condamné Jean-Marie Villemin le 16 décembre 1993?*" ["*What was the sentence received by Jean-Marie Villemin on 16 December 1993 ?*"]), Google Desktop does not find any answer because the acceptable answers "cinq ans de prison" (five years of prison) or "cinq années d'emprisonnement" (a five-year prison sentence) does not contain, in French, any word of the query. Similarly the search engine has many difficulties to display the development of acronyms like those of the question 141 ("*Que signifie CGT?*" ["*What is the significance of CGT?*"]), or of the question 319 ("*Qu'est-ce que l'EEE?*" ["*What is EEE?*"]). Because the answers are developments of these capital letters which are not so frequent, except when the acronym is rare, as in this case the acronym is often followed or preceded by his significance, like for the questions 327 ("*Qu'est-ce que le Cermoc ?*" ["*What is Cermoc?*"]) or 330 ("*Qu'est-ce que l'OACI ?*" ["*What is OACI?*"]). For these two questions, Google Desktop returns the correct answer in the first snippet.

## 5 Perspectives

The evaluation was designed in such fair way to take into account all the differences between the Information Retrieval engines and the Question Answering systems.

We made all possible efforts not to favour the QA systems and avoid non equitable comparison. For example, the evaluation includes the requests to Google Desktop made with the most sophisticated achievable query syntax to generate a return of the best answers, knowing that if they were keyed in as for the natural language requests, their success rate would have dropped considerably (see Table 1).

It is most unlikely that one is able to formulate queries in Boolean like style as quickly as in NL questions. Conversely, for a same given number of characters, reading Google snippets requires most likely far more time than reading complete NL sentences. However, despite all these metrical choices more favourable to the classical search engine, the Question Answering system obtains better results with regard to the quality of the answers and to the user effort.

If we were able to compare the Web versions of Google and Qristal, the results would be probably different.

First, because Qristal uses the search engines as a meta-engine without any indexation. Next, because GoogleWeb is really faster at displaying the results from the Web than Google Desktop. At last because the redundancy, due to the large volume of indexed pages on the Web, allows the implementation of some very successful techniques.

For example it may happen that you find the questions in natural language followed by their answers inside Web pages and this in such a way that asking a request in natural language in Google, you can obtain sometimes a very pertinent answer. To the question "*Pourquoi le ciel est bleu?*" ("*Why the sky is blue?*") or to the question "*Pourquoi la mer est bleue?*" ("*Why the sea is blue?*"), Google Web returns in first rank snippets and documents very accurately.

However the analysis of the documents and contained answers permit to the Question Answering systems to return more accurate answers. For example, with the request "*capitale anglaise*" ("*English capital*"), Google returns a lot of snippets containing the phrase "*capitale anglaise*" ("*English capital*") but not the word Londres or London in these snippets. In an Information Retrieval engine the answers are very often less justified by the context than it is the case with Question Answering systems. This is because the snippets group essentially words contained in the query. For example, to the question 26 ("*Qui a écrit Germinal?*" ["*Who wrote Germinal?*"]), converted in Google syntax by : "*auteur Germinal*" ["*writer Germinal*"]), the search engine returns "Émile Zola" in the second snippet but the snippet "*L'exposition "Emile Zola, photographe" fait escale*" ("*The exhibition "Emile Zola, photograph" stops at*) would be considered as an answer out of its context and non receivable within a campaign like TREC, even if we can read in the text : "*L'auteur de "Germinal", l'écrivain français Emile Zola (1840-1902), était aussi un photographe de talent*" ("*The author of Germinal, the writer Emile Zola (1840-1902), was also a talented photograph*"). We could almost say that the classical search engines return far better results when the user already knows the answer to his query.

A complete compared benchmark and exhaustive evaluation of search engines and question answering systems needs to be made on the Web.

The evaluation method described above could be applied but, knowing the difficulty to validate a Web corpus of answers, specially the difficulty to keep it referentially constant, the effort to estimate the quality of the returned answers would be far much enormous than the one engaged for this evaluation on a closed corpus.

## 6 Conclusion

We described a metrical method to compare Question Answering systems and Information Retrieval engines on a hard disk-based corpus. Applied to our system Qristal and to the search engine Google Desktop, this method shows that the improvements due to question answering systems, especially in terms of user effort, are both qualitative and quantitative. Taking into account or not the query keying time, the question answering system is 2 to 6 times faster than the Information Retrieval engine.

This evaluation, made almost automatically with the corpus EQueR, focuses mainly on factual or definition types of requests which defines the majority of the requests concerned in the QA system evaluation campaigns. On more complex questions, like those beginning by "comment" ("how..."), the QA systems obtained less satisfying results than for factual questions but, on this type of questions, the search engines like Google proved to be also less accurate. A more exhaustive and thorough study on these types of requests or on questions beginning by "pourquoi" ("why") would possibly confirm these results, although here only a few questions were of these types.

At last, a similar metrics and applied evaluation remain to be endeavoured on a web-based corpus despite the entailing difficulties.

## References

- Amaral C., Laurent D., Martins A., Mendes A., Pinto C. (2004), Design & Implementation of a Semantic Search Engine for Portuguese, In Proceedings of the Fourth Conference on Language Resources and Evaluation.
- Amaral C., Figueira H., Martins A., Mendes A., Mendes P., Pinto C., (2005) Priberam's Question Answering System for Portuguese. In Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria
- Ayache C., Choukri K., Grau B. (2004) Campagne EVALDA/EQueR Évaluation en Question-Réponse.
- [http://www.technolanguen.net/IMG/pdf/rapport\\_EQUER\\_1.2.pdf](http://www.technolanguen.net/IMG/pdf/rapport_EQUER_1.2.pdf)
- Ayache C., Grau B., Vilnat A., (2005) Campagne d'évaluation EQueR-EVALDA : Évaluation en question-réponse. In TALN & RECITAL 2005, Tome 2 - Ateliers et tutoriels, pp. 63-72
- Buchholz S., (2002) Open-Domain Question Answering on the World Wide Web. Tutorial, <http://tcc.itc.it/research/textec/topics/question-answering/Tut-Bucholtz.html>.
- Kwok C., Etzioni O., Weld D.S., (2001) Scaling Question Answering to the Web. In Proceedings International WWW Conference(10), Hong-Kong
- Laurent D., Varone M., Amaral C., Fuglewicz P. (2004) Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies, First International Workshop on Proofing Tools and Language Technologies, Patras, Grèce.
- Laurent D., Séguéla P., Nègre S., (2005-1) QRISTAL, système de Questions-Réponses. In TALN & RECITAL 2005, Tome 1 - Conférences principales, pp. 53-62
- Laurent D., Séguéla P., Nègre S., (2005-2) Cross-Lingual Question Answering using Qristal for CLEF 2005. In Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria
- McGowan K., (2005) Emma : A Natural Language Question Answering System for umich.edu. <http://www-personal.umich.edu/~clunis/emma/>
- Radev D.R., Qi H., Wu H., Fan W., (2002) Evaluating Web-based Question Answering Systems. In the Proceedings of the 11th WWW conference, Hawaii.
- Richaudeau F., Gauquelin M. et F., (1977) La méthode complète de lecture rapide "Richaudeau", éd. Retz-CEPL, Paris.
- Vallin A. Giampiccolo D., AunimoL., Ayache C., Osenova P., Peñas A., de Rijke M., Sacaleanu B., Santos D., Sutcliffe R. (2005) Overview of the CLEF 2005 Multilingual Question Answering Track. In Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria
- Voorhees E.M., (2005) Overview of the TREC 2004 Question Answering Track. In Proceedings of the Thirteenth Text REtrieval Conference (TREC 2004).