

## TechPak Form

**Welcome** to the *yet2.com* TechPak form (Copyright © 2001 yet2.com Inc. All rights reserved.) Using this form, you can create, edit, and submit a TechPak for the *yet2.com* Web site.

The form assembles the information that a person most needs to make a licensing decision about your technology. As you complete the form, think of it as a new kind of technical literature—one designed to display your technology, clearly explain its nature and its benefits, and suggest applications where others can make good use of it.

You already may be familiar with patents. A patent is designed to claim much and reveal as little as possible. A TechPak goes much further by providing, for example, a description of your technology's functionality in terms understandable by people not in your discipline. You'll need to put yourself in the shoes of people who are looking for technical solutions, but who may not yet be looking for *your* solution.

This document includes the following sections:

1. Technology Overview
2. Novelty
3. Development Status and Potential Applications
4. Intellectual Property
5. Provider Company Demographics
6. Technical Details
7. Licensing
8. Internal Use Only

### **Getting Help**

There are several places to get help with this form and with TechPaks in general:

- **Click the Help button** next to a field for information specific to that field. Help includes an explanation of the field, tips, and an example of the kind of information the field holds. To display this information, you will need an Internet connection. This form will contact Help on the *yet2.com* Web site.
- **Contact Customer Support** at *yet2.com*. In the U.S., our **toll-free** support number is 1-877-885-9075. Internationally, you can reach Customer Support at 1-617-557-3800. Our fax number is 1-617-523-8232. You can email Customer Support at **support@yet2.com**.

*yet2.com* also can help you assess your technology portfolio, determine which of your technologies may obtain the most results on the *yet2.com* Web site, and write TechPaks for you. For more information on our services, please contact **sales@yet2.com**.

Spell Check Form Fields

## 1. Technology Overview

---

### Title (required)

?

Question-Answering System in Natural Language on Internet and Intranets  
Malek Boualem, Pascal Filoche, France Telecom

### Promotional Description (required)

?

This technology developer has devised a Question-Answering System that outperforms conventional search engines on the Internet. It allows the user who wants to get particular information to ask a question in natural language and to obtain a concise up-to-date answer. The system analyses the enquirer's question and the Web pages identified by standard search engines to find the actual answer, for instance, the question "Who is the president of the USA ?" will receive in return the precise answer : "George W. Bush". The system is designed to work in most languages and is currently customized for English, French, Polish, Spanish, German and Arabic.

### Insert Image (optional)



Click button to insert image. Your image may be slightly distorted when placed in this form, but it will appear correctly in the final display.

insert image...

remove image...

### Technology Type (optional; select all that apply)

?

Process  Design  Material  Software

### Display Company Identity

?

Display Company Logo

### Technology Benefits Summary (recommended)

?

## 1. Technology Overview

---

This system is particularly suitable for users seeking precise answers to factual questions. The user does not have to read all the documents provided by a search engine in order to find the particular answer requested.

## 2. Novelty

---

### Technology Benefits (required)

?

Start Technology Benefits

The Question-Answering System can :

- Retrieve answers to questions where conventional search engines only retrieve documents.
- Search through either the Internet, Intranets and can be integrated to any search or meta-search engine.
- Provide a convenient route direct to the answer and return up-to-date information.
- Thanks to natural language processing, select the documents containing the answers more efficiently.

Finish Technology Benefits

### Technology Differentiation and Uniqueness (recommended)

?

Start Technology Differentiation and Uniqueness

The natural language processing technology imbedded in the Question-Answering System differentiates this search facility from all the competition and provides the following unique features :

- Directly retrieves the answers, contrary to other natural language retrieval systems, which only display text passages likely to contain the answers.
- Provides answers to any kind of factual questions, as it is not restricted to a closed pre-formatted database to search for the answer.
- Searches for answers in various forms by using morphological features and provides synthetic answers.
- Multilingual including various language families such as Indo-European or Semitic languages.
- Customizable according to users' needs.

Finish Technology Differentiation and Uniqueness

### 3. Development Status and Potential Applications

---

#### Standard Listing

Click one of the following buttons to change the Listing Type:

?

Standard Listing

Life Sciences Listing

#### Development Stage (recommended)

?

Commercialization

#### Application and Potential Advantage (recommended)

?

Added-value with Internet and Intranet search engines	<a href="#">delete</a>
Internal data and document management systems	<a href="#">delete</a>
Remote provision of data and information	<a href="#">delete</a>
Business and strategic information surveys	<a href="#">delete</a>

add row...

#### Development Status Summary (recommended)

?

The Question-Answering service is ready to be integrated and has been demonstrated by the technology provider as being extremely beneficial to information retrieval applications.

## 4. Intellectual Property

---

Specific Patent Information (optional)			?
Patent Number	Country	Status	

add row...

### Patent Summary

Generate Summary

Input Summary

**Generate Summary**

A summary of the specific patent information entered above will be generated for you after your form is submitted and processed. If you would like to enter your own summary information, please click the Input Summary button above.

## 5. Provider Company Demographics

---

### Related Competency (optional)

?

This Question-Answering system is part of a set of innovative natural language processing software that are also available :

- automatic text summarization,
- thematic classification of documents,
- language and character encoding identification,
- document indexation,
- analysis of web sites,
- software and service linguistic localization.

## 6. Technical Details

---

Long Description (optional)

?

Start Long Description

### **Introduction**

The Internet has provided the world wide facility for rapid access to information. Most of this information is retrieved through search engines using keywords. However, most of the time, the user can only enter keywords into the search engine, so that the precise information he is looking for does not appear explicitly. The user then has to sift through the references to find out the real information wanted.

An example of this follows with the question "Who is the president of the USA?". If the user inputs "president" and "USA" into a search engine, the latter will return all the documents containing these words. One of these references could be "*Society for mathematical and computational aesthetics – Society president : Michael Leyton (USA)*". The document may contain all the keywords in entry, but George W. Bush's name does not appear. Even if the user is more experimented with search engines and inputs the query "president of the USA", the search engine may not take into account the words "of" and "the". In this case, the results of the search are once again not satisfying. This demonstrates two major drawbacks of traditional search engines. The first one is that the answer is lost in a huge number of documents and a lot of time may be required to read through every document to find out the relevant information. The second drawback is that, very often, the contents of the documents bear little resemblance to those expected.

In order to retrieve a specific answer, the search facility has to interpret the semantics and morphology in a question to provide the answer. For example, if the user wants to know how high the Eiffel Tower is, he first needs to formulate the right keywords for the search engine so as to avoid to get documents concerning the Pisa Tower or Gustave Eiffel's life. The following picture shows the process of sending the query "high Eiffel Tower" to a standard search engine. Not only is the number of results considerable but the user is also obliged to read through every document to find out the answer to his question.



## 6. Technical Details

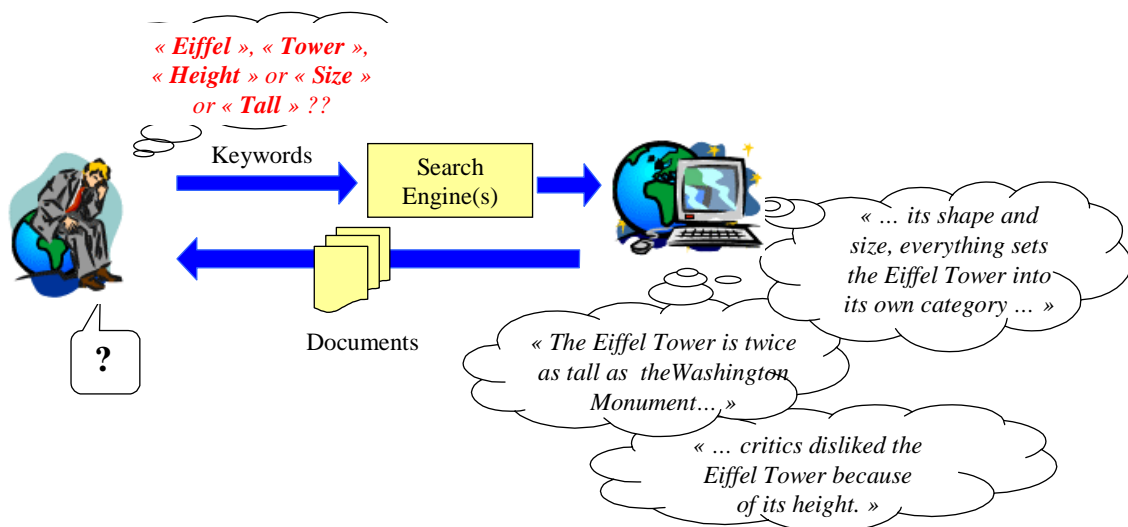


Fig.1 Searching the Internet using only search engines

### The Question-Answering System (Q&A)

By adding natural language processing technologies and extending the linguistic coverage, the user can ask his question and obtain a concise up-to-date answer in a short time.

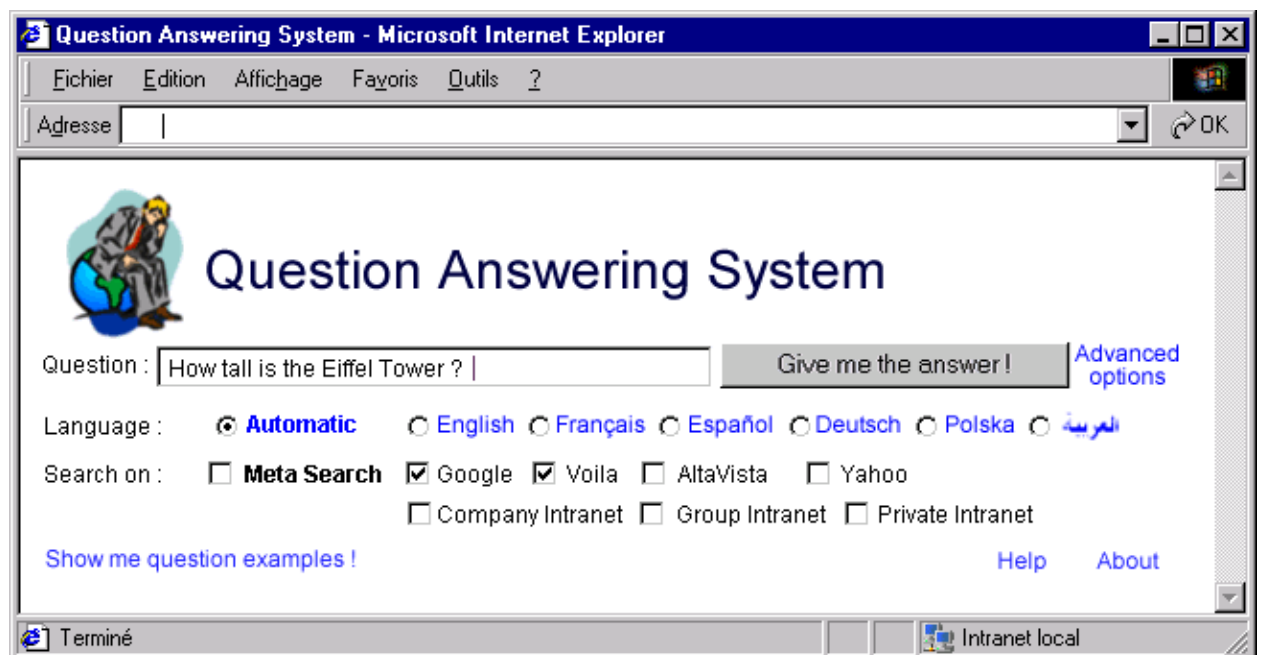


Fig.2 Stand-alone Question-Answering interface

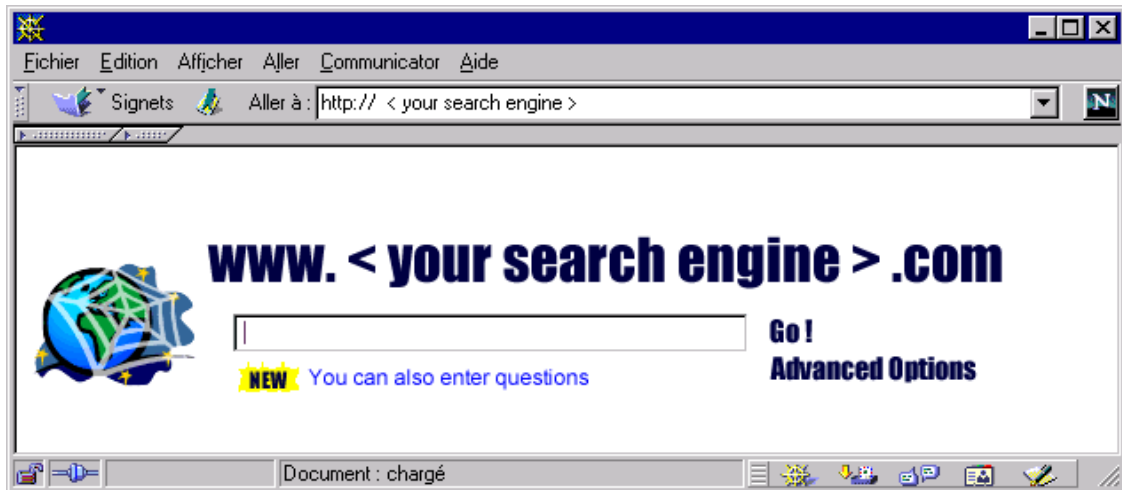
The user asks a question to the Question-Answering system in everyday language. He may then select a specific language or let the system identify the language automatically

## 6. Technical Details

---

when processing his question. He may also select which search engines should be used (on the web or on his intranet), or let the system use them all.

Moreover the Question-Answering can be integrated more intimately within an existing search engine architecture. The user still can use the usual interface of the search engine, and can also enter questions rather than only keywords.



*Fig.3 Search engine interface powered by Question-Answering functionalities*

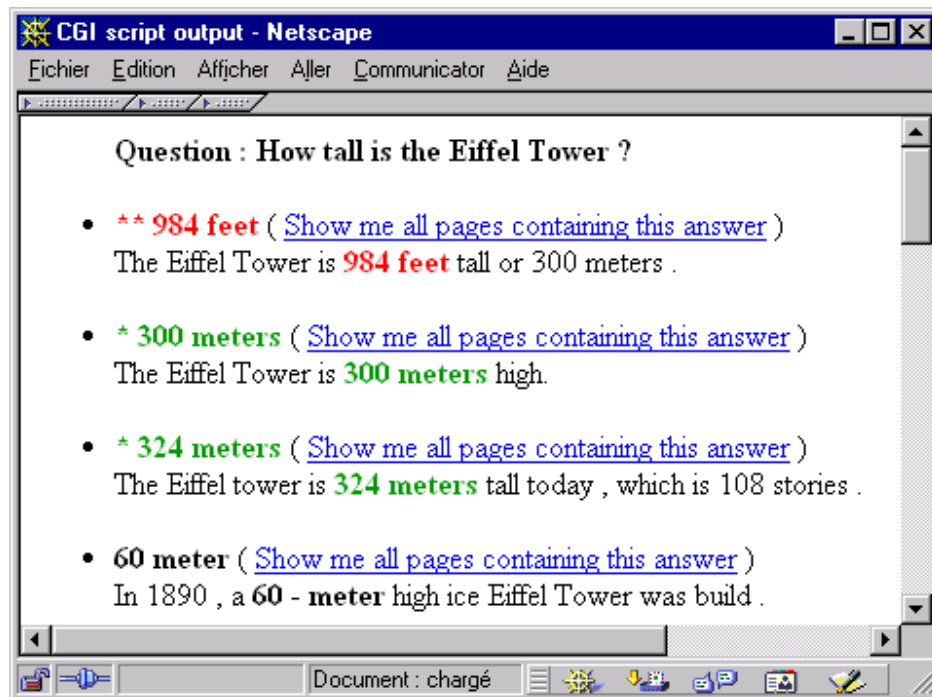
The Question-Answering system can process many different question types in several languages. Here are some examples of questions that can be asked in English :

- |  |
|--|
| <p>Questions about people :</p> <ul style="list-style-type: none"><li>• Who built the Eiffel Tower ?</li><li>• Who wrote Moby Dick ?</li></ul> <p>Questions about companies, etc. :</p> <ul style="list-style-type: none"><li>• Which are the Japanese mobile operators ?</li><li>• Which are the subsidiaries of France Telecom ?</li></ul> <p>Questions about acronyms :</p> <ul style="list-style-type: none"><li>• What does VXML mean ?</li><li>• What does ADSL mean ?</li></ul> <p>Questions about places :</p> <ul style="list-style-type: none"><li>• Where did Gandhi die ?</li><li>• Where is Montmartre ?</li></ul> <p>Questions about dates of events :</p> <ul style="list-style-type: none"><li>• When did Colombus discover America ?</li><li>• When was Gustave Eiffel born ?</li></ul> <p>Questions about quantities :</p> <ul style="list-style-type: none"><li>• How high is the Eiffel Tower ?</li><li>• For how much did France Telecom buy Orange ?</li></ul> <p>Questions about cause, effect, reasons, processes, etc. :</p> <ul style="list-style-type: none"><li>• How is couscous cooked ?</li><li>• Why do tourists visit Paris ?</li></ul> |
|--|

## 6. Technical Details

---

Once the user has asked his question, the system sends several queries to the specified search engines, using different sets of keywords generated from the user's question (using morphological and semantic relations – e.g. searching “buys” as well as “bought”, and “high” as well as “tall”). While doing this, the system crawls the documents selected by the engines, and extracts possible answers to the question. Syntactic variations of the answers are taken into account by the system (e.g. “Melville wrote Moby Dick” / “Moby Dick was written by Melville”). Eventually, those answers are ranked, and displayed to the user with context sentences found during the process :



*Fig.4 Answers are ranked and displayed to the user*

The fact that only a few words are dynamically returned to answer a question and that they are extracted from the current image of the Internet or Intranet makes this system unique, thanks to our know-how in natural language processing.

### **General description of the Question-Answering System**

This system is particularly suitable for users seeking precise answers to factual questions. For these kinds of queries the main drawbacks of the search engines can be avoided. The user can precisely define his information need through the use of his natural language avoiding many irrelevant documents.

## 6. Technical Details

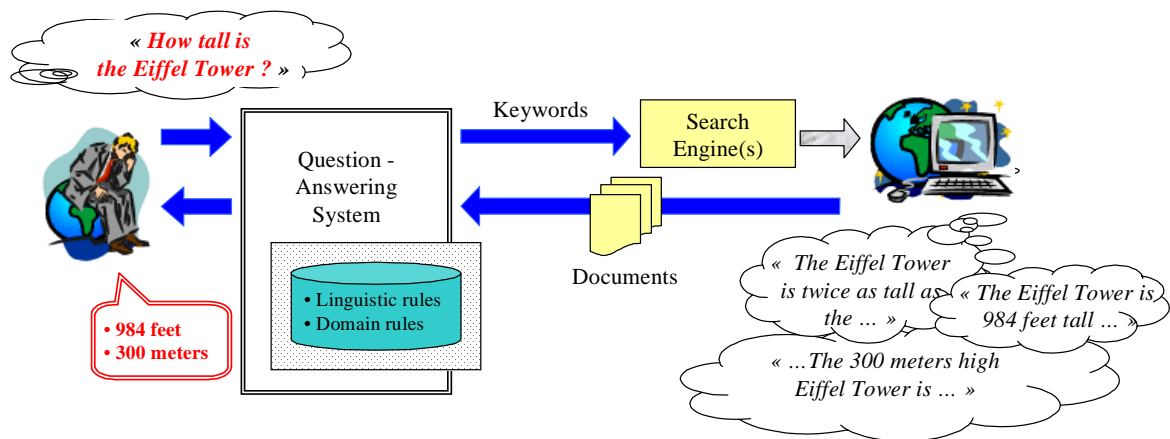


Fig.5 Searching the Internet and intranets using Q&A on search engines

Given a question expressed in natural language, the system produces one or several relevant set(s) of keywords which are then sent to a traditional “keyword input” search engine. The documents whose URLs are returned by the search engine are downloaded and analyzed to find out whether or not they contain sentences likely to answer the question (i.e. containing all the selected keywords). The answer is extracted as a chunk of text and is done by building abstract syntactic representations (“patterns”) of the possible answers.

The underlying mechanism keeps track of the different steps during processing. This allows for the visualization of either the answer or the overall context such as the sentence embedding the answer as a preview, the parsed HTML page and the URL.

*This system can be seen as a linguistic filter plugged on a search engine which allows for big profit in terms of answer precision.*

The linguistic analysis is based on lexical access, morphological analysis and shallow parsing (“chunking grammar”). The result of this analysis consists of lemmas, a set of features including parts of speech and usual grammatical categories (such as tense, gender, number, person), morpho-semantic features for proper nouns, numbers or places and syntactic tags associated to the various nominal, adjectival and verbal phrases recognized by the chunking grammar. This analysis applies to both the question and to a set of candidate sentences for the answer.

*Question-Answering is strongly related to textual information extraction, an increasingly important field of natural language processing, which is one of the components of this system.*

### **Multilingual Question-Answering System**

The system is multilingual. On the one hand this means that designing a Question-Answering system in a new language only involves the development of linguistic resources for this language and a set of specific rules leading to the extraction patterns.

## 6. Technical Details

---

On the other hand, the system will integrate cross-language information retrieval functionalities that allow it to retrieve answers even if they are not in the user's language.

Finish Long Description

### Frequently Asked Questions (optional)



add row...

## 7. Licensing

---

<b>Type of Collaboration Sought (recommended; select all that apply)</b>	?
<input type="checkbox"/> Venture Funding <input type="checkbox"/> R&D Contract <input type="checkbox"/> Joint Venture <input checked="" type="checkbox"/> License <input type="checkbox"/> For Sale	

<b>Collaboration Description (optional)</b>	?

<b>Terms &amp; Restrictions (optional)</b>	?
No restrictions	

<b>Seller Support (recommended; select all that apply)</b>	?
<input checked="" type="checkbox"/> Documentation <input type="checkbox"/> Personnel <input type="checkbox"/> Third Party	

<b>Available Technical Assistance (recommended)</b>	?
Negotiable	

## 8. Internal Use Only

---

<b>Internal Reference Code (optional)</b>	?
FranceTelecom-S26-02-10/5D : QAS	

<b>Primary Author (optional)</b>	?
Name	John Moreland
Phone	+44 (0) 1628 816585
Email	john.moreland@jratech.co.uk

<b>Primary Inventor (optional)</b>	?
Name	
Phone	
Email	

<b>Primary Licensing Manager (optional)</b>	?
Name	
Phone	
Email	

<b>Notes (optional)</b>	?
The France Telecom R&D logo will be inserted by Yet2.com when this TechPak is loaded on to the website.	

Form Version: 12.1

Copyright © 2001 yet2.com Inc. All rights reserved