

Apport de la syntaxe dans un système de question-réponse : étude du système FIDJI.

Véronique Moriceau Xavier Tannier
Université Paris-Sud 11 / LIMSI-CNRS
BP 133, 91403 Orsay Cedex, France
{moriceau, xtannier}@limsi.fr

Résumé. Cet article présente une série d'évaluations visant à étudier l'apport d'une analyse syntaxique robuste des questions et des documents dans un système de questions-réponses. Ces évaluations ont été effectuées sur le système FIDJI, qui utilise à la fois des informations syntaxiques et des techniques plus "traditionnelles".
La sélection des documents, l'extraction de la réponse ainsi que le comportement selon les différents types de questions ont été étudiés.

Abstract. This paper presents some experiments aiming at estimating the contribution of a syntactic parser on both questions and documents in a question-answering system. This evaluation has been performed with the system FIDJI, which makes use of both syntactic information and more "traditional" techniques.
Document selection, answer extraction as well as system behaviour on different types of questions have been experimented.

Mots-clés : Systèmes de questions-réponses, analyse syntaxique, évaluation.

Keywords: Question-answering, syntactic analysis, evaluation.

1 Introduction

Nous présentons ici les résultats d'une série d'évaluations sur le système FIDJI¹ (Finding In Documents Justifications and Inferences), un système de questions-réponses en domaine ouvert pour le français, utilisant notamment l'analyse syntaxique des textes et des questions.

Les objectifs de cet article sont les suivants :

- Montrer que la syntaxe, exploitée en complément de techniques plus "traditionnelles" de questions-réponses, peut conduire à une amélioration notable des résultats globaux ;
- Étudier à quelles étapes du traitement et pour quels types de questions l'utilisation de l'analyse syntaxique peut apporter une réelle plus-value.

Dans un premier temps, la section 2 résume l'architecture et le fonctionnement du système FIDJI. Puis l'évaluation modulaire, cœur de cet article, est présentée à la section 3, avec une présentation globale des résultats, puis des études sur l'apport de l'analyse syntaxique lors de

¹Ce travail est financé partiellement par le projet CONIQUE ANR-05-BLAN-008501 et par OSEO dans le cadre du programme Quaero.

la phase de sélection des documents, de l'extraction de la réponse, et enfin sur les différences entre les différents types de questions.

2 Une approche syntaxique pour la validation de réponses

2.1 Présentation de FIDJI

La plupart des systèmes de questions-réponses peut extraire une réponse à une question factuelle quand celle-ci est explicitement présente dans le texte, mais ils ne sont pas capables de combiner plusieurs informations pour produire une réponse.

Le système de questions-réponses FIDJI vise à introduire des mécanismes de compréhension reposant sur des inférences. L'objectif est de produire des réponses qui sont entièrement validées par des extraits de textes (des passages).

Pour la première étape de développement de ce système, nous avons introduit l'utilisation d'un analyseur syntaxique robuste, d'une part pour le traitement et la décomposition des questions et d'autre part pour l'analyse des documents.

L'ajout d'informations syntaxiques a plusieurs buts :

- Faciliter l'obtention des caractéristiques portées par la question : type de la question, de la réponse attendue, entités nommées, etc.
- Décomposer éventuellement la question avec pour objectif de valider une réponse potentielle dans plusieurs documents.
- Sélectionner des passages candidats (*i.e.* contenant peut-être la réponse) de façon plus précise, en utilisant non plus des mots-clés simples, mais des dépendances syntaxiques.
- Enfin, extraire la réponse de ces passages à partir de la combinaison des informations issues de la question et des phrases étudiées.

Beaucoup de systèmes de questions-réponses utilisent des informations syntaxiques, en particulier les relations de dépendance, principalement pour l'extraction des réponses. Deux approches émergent : la première consiste à rechercher un appariement exact entre les relations de dépendance de la question et celles d'un passage (Katz & Lin, 2003), tandis que la seconde approche calcule une distance d'édition entre les arbres représentant la question et le passage (Ligozat, 2007). En ce qui concerne la décomposition des questions, les stratégies proposées se positionnent à un niveau à la fois syntaxique et sémantique (Katz *et al.*, 2005; Hartrumpf *et al.*, 2008).

Notre but est d'extraire et de valider des réponses en allant au-delà d'un appariement syntaxique exact entre la question et le passage, et cela sans utiliser de ressources sémantiques. Dans ce contexte de validation de réponse, nous avons remarqué que la stratégie de validation à appliquer (validation par un seul ou plusieurs documents) peut être guidée par la question, et en particulier par le type de réponse attendu.

Une présentation détaillée du fonctionnement de FIDJI et de chacun de ses modules est donnée dans (Moriceau *et al.*, 2009).

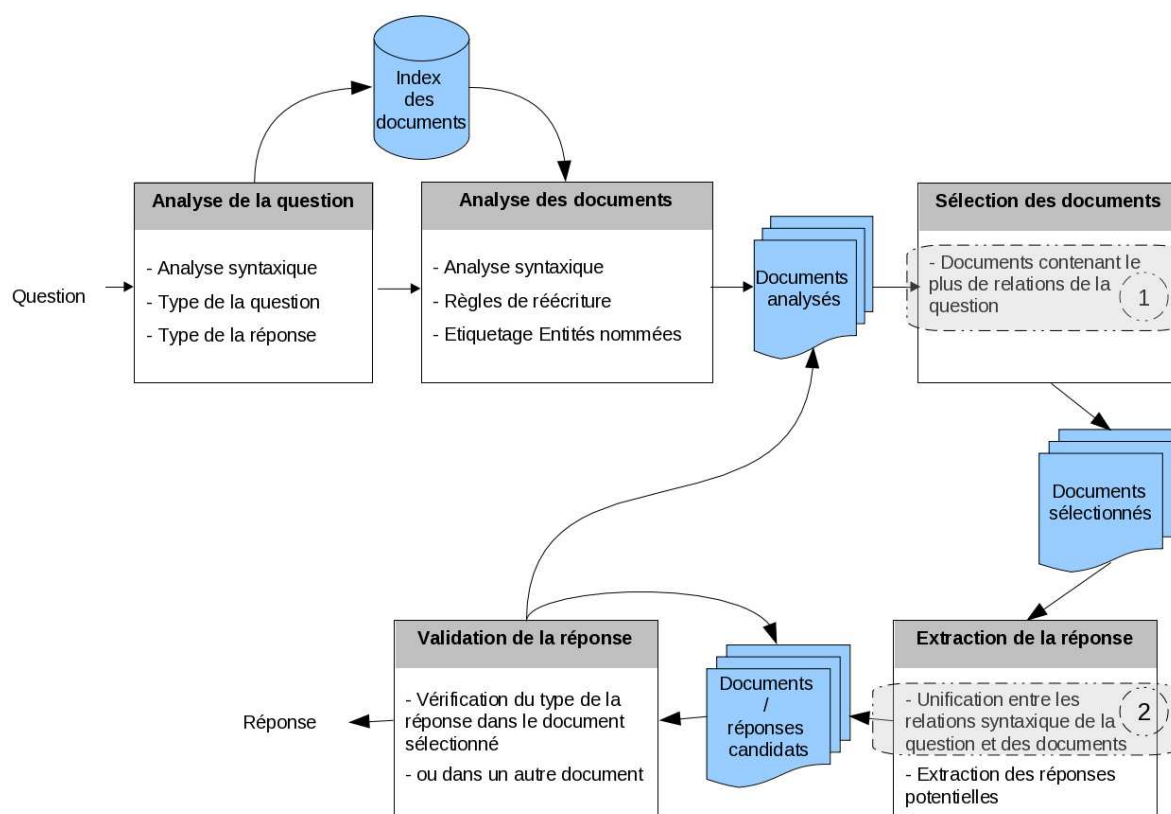


FIG. 1 – Architecture de FIDJI

2.2 Principe général

Nous présentons ici l'architecture générale ainsi que deux exemples : 1) l'extraction d'une réponse et la validation de son type dans le document sélectionné, et 2) l'extraction d'une réponse et la validation de son type dans plusieurs documents.

La figure 1 présente l'architecture de FIDJI. L'ensemble des documents est indexé par le moteur de recherche Lucene (Hatcher & Gospodnetić, 2004). On extrait de la question les relations syntaxiques à rechercher, le type de la question (liste, définition, fait...) ainsi que le type de la réponse (type d'entité nommée ou type plus spécifique – voir l'exemple 2) et les mots-clés significatifs². L'index est interrogé avec ces mots-clés et les 100 premiers documents sont conservés.

L'approche consiste à déterminer, pour une question donnée, si toutes les caractéristiques de la question (en l'occurrence les dépendances syntaxiques) peuvent être trouvées dans un ou plusieurs documents. Notre système s'appuie sur les analyses syntaxiques produites par SynTex (Bourigault & Fabre, 2000), un analyseur robuste pour le français.

2.2.1 Exemple 1

Dans un premier temps, l'analyse de la question fournit les relations de dépendance syntaxique, le type de la question et le type de réponse attendu. Par exemple :

²C'est-à-dire les mots étiquetés comme nom, verbe, adjectif ou adverbe par l'analyseur syntaxique.

Question : Quelle ville a été secouée par un tremblement de terre le 17 janvier ?

Dépendances : DATE(17 janvier)
attribut_de(tremblement, terre)
SUJ(secouer, REPONSE)
AUX(être, secouer)
modif_par(secouer, tremblement)

Type de la question : factuelle

Type de réponse attendu : ville (entité nommée)

Les phrases contenant le plus de relations en commun avec la question sont ensuite sélectionnées dans un sous-ensemble de documents de la collection. Par exemple :

Passage : Le tremblement de terre qui a secoué, <date>lundi 17 janvier</date>, <ville>Los Angeles</ville> ne serait pas associé directement à la fameuse faille...

DATE(17 janvier)
attribut_de(tremblement, terre)
SUJ(secouer, tremblement)
OBJ(secouer, Los Angeles)
SUJ(secouer, Los Angeles)
AUX(être, secouer)
modif_par(secouer, tremblement)
...

Ici, toutes les relations de la question sont présentes dans l'analyse du passage (on note qu'ici, les relations en italique ont été obtenues par une règle de réécriture de la voix active vers la voix passive). Par unification avec la variable réponse de la question, la réponse potentielle qui est extraite est donc *Los Angeles*. Il reste alors à vérifier que le type de cette réponse correspond bien au type de réponse attendu (*ville*).

Pour cela, les entités nommées des documents ont été préalablement étiquetées en utilisant un ensemble d'environ 160 types (Rosset *et al.*, 2007) (*e.g.* personne, organisation, lieu, nationalité, date, nombre, etc.). Ainsi, le passage étiqueté en entités nommées permet de valider le type de la réponse candidate extraite (*Los Angeles*).

2.2.2 Exemple 2

Soit la question : Quel premier ministre s'est suicidé en 1993 ?

Dépendances :

DATE(se suicider, 1993)
SUJ(se suicider, REPONSE)
attribut(REPONSE, ministre)
attribut(ministre, premier)

Type de la question : factuelle

Type de réponse attendu : personne (type précis : premier ministre)

Et la phrase sélectionnée suivante :

<personne>Pierre Bérégovoy</personne> s'est suicidé en 1993.

DATE(se suicider, 1993)

SUJ(se suicider, Pierre Bérégovoy)

La variable REPONSE s'unifie avec *Pierre Bérégovoy* qui devient une réponse candidate : le type d'entité nommée et les deux premières dépendances de la question sont ainsi vérifiés dans cette phrase. Il manque les deux dépendances suivantes, concernant le type précis de la réponse (Pierre Bérégovoy était-il premier ministre ?).

Une nouvelle question est donc construite pour valider la réponse candidate. La validation est opérée en deux étapes. Tout d'abord, le système vérifie que la réponse candidate est bien un *ministre*, en recherchant une relation 'attribut' (`attribut(Bérégovoy, ministre)`) dans l'ensemble des documents. Si cela est confirmé, le type étendu est également vérifié et les deux relations `attribut(Bérégovoy, ministre)` et `attribut(français, ministre)` sont attendues dans la même phrase.

Dans les cas où la variable REPONSE n'existe pas ou n'est pas instanciée par des relations du passage, le système recherche un ou plusieurs candidats dans la phrase parmi les entités nommées du type attendu.

3 Apport de la syntaxe dans FIDJI

Comme nous l'avons indiqué, si les résultats généraux ont de l'intérêt, nous nous sommes surtout attachés à l'estimation du rôle de la phase d'analyse syntaxique dans notre système. En effet, il est important d'évaluer le bénéfice de la syntaxe par rapport aux techniques plus classiques, et ce indépendamment des autres caractéristiques propres à un système donné.

Plusieurs points sont ici évalués :

- les performances d'une approche syntaxique associée à des techniques classiques en question-réponse,
- l'apport de la stratégie de décomposition des questions,
- l'apport de la syntaxe pour la sélection des documents,
- l'apport de la syntaxe pour l'extraction des réponses,
- les différences concernant l'extraction des réponses selon le type de la question.

Nous avons évalué le système sur les données de test de la campagne d'évaluation CLEF 2005 (Vallin *et al.*, 2005). La collection de documents est composée d'environ 177000 articles de journaux en français (Le Monde et ATS 1994-1995 (environ 2 Go)). Ces documents sont censés être syntaxiquement corrects. Les questions sont factuelles ou de type définition. Lors de cette campagne, les systèmes sont autorisés à proposer 3 réponses pour chaque question, ces réponses devant être classées par ordre de confiance.

La méthodologie employée et les résultats obtenus sont décrits dans les sections suivantes.

3.1 Performances globales de FIDJI

Le tableau 1 présente les résultats obtenus par FIDJI sur les données de CLEF 2005 (nombre de réponses correctes proposées en première position). Nous comparons ces résultats avec ceux de QRISTAL (Laurent *et al.*, 2005; Laurent *et al.*, 2006), le meilleur système de questions-réponses pour le français lors de ces campagnes.

| Type de question | FIDJI | QRISTAL |
|------------------|--------|---------|
| Factuelle | 53 % | 59 % |
| Définition | 78 % | 86 % |
| TOTAL | 58.5 % | 64 % |

TAB. 1 – Résultats de FIDJI sur les données de CLEF 2005.

Le nombre de réponses correctes s'élève à 58.5 % pour les réponses données en rang 1 et à 70 % si l'on considère les rangs 1 à 3.

Les résultats de FIDJI sont inférieurs à ceux de QRISTAL qui utilise aussi une approche syntaxique mais qui bénéficie de l'utilisation de nombreuses ressources telles que des dictionnaires et des ontologies. Cependant, notre système se place "virtuellement" à la deuxième place puisque le système qui a atteint la deuxième place lors de CLEF 2005 a obtenu un score de 35 % de réponses correctes.

Nous évaluons maintenant l'apport de la stratégie de décomposition syntaxique des questions dans le but de valider une réponse grâce à plusieurs passages.

Dans l'ensemble de questions de CLEF 2005, sur les 51 questions qui peuvent être décomposées en sous-questions en appliquant notre stratégie, FIDJI trouve une réponse correcte (sans tenir compte du rang) pour 32 questions. Rappelons que chaque fois que la décomposition des questions est appliquée, le système peut rechercher des justifications à la réponse dans des documents différents. Parmi ces 32 réponses correctes, 22 % ont une justification dans plusieurs documents. En revanche, si FIDJI n'utilise pas la stratégie de décomposition et de validation par plusieurs documents, seulement 64 % des réponses sont correctes (au lieu de 70 %).

Le tableau 2 montre le nombre de réponses correctes dont la justification a été trouvée dans un ou plusieurs documents.

| Réponses correctes en rang 1 à 3 | CLEF 2005 |
|---|------------------|
| Avec justification dans 1 document | 78 % |
| Avec justification dans plusieurs documents | 22 % |
| Sans validation multi-documents | 64 % |
| Avec validation multi-documents | 70 % |

TAB. 2 – Evaluation de la stratégie de décomposition des questions.

Ces résultats montrent une amélioration des performances du système lorsque l'on utilise une stratégie de décomposition des questions afin de valider les réponses par l'intermédiaire de plusieurs documents (Moriceau *et al.*, 2009).

3.2 Apport de la syntaxe pour la sélection des passages

Dans les systèmes de questions-réponses, la sélection des passages candidats est généralement effectuée en prenant en compte les mots de la question. Dans FIDJI, nous nous situons au niveau des dépendances syntaxiques, en sélectionnant les phrases comportant le plus de dépendances (nom de la relation et arguments) en commun avec la question.

Pour évaluer l'apport de cette technique, nous l'avons comparée avec une sélection des phrases comportant simplement le maximum de mots-clés de la question (désactivation du module noté ① dans la figure 1).

L'évaluation est effectuée sur la collection de CLEF 2005 et les résultats sont reproduits dans le tableau 3. La progression du nombre de bons passages (passages contenant la réponse en rang 1) est tout à fait significative (+8,1 %). Mais le nombre de bonnes réponses en rang 1 augmente encore plus (+15,8 %). Ainsi, la syntaxe ne permet pas seulement de récupérer plus de bons passages candidats, mais ces passages sont même plus pertinents pour une extraction efficace de la réponse.

Par ailleurs, on constate que le progrès pour les rangs 1 à 3 est plus faible. Cela semble montrer que les passages renvoyés sont comparables, mais que l'utilisation des relations syntaxiques en améliore le classement.

| | Mots-clés | Dépendances |
|---|-----------|-------------|
| Réponses correctes en rang 1 | 50.5 % | 58.5 % |
| Réponses correctes en rang 1 à 3 | 68.5 % | 70 % |
| Passages contenant la réponse en rang 1 | 62.6 % | 67.7 % |
| Passages contenant la réponse en rang 1 à 3 | 72.7 % | 75.8 % |

TAB. 3 – Apport de l'utilisation de la syntaxe pour la sélection des passages candidats.

3.3 Apport de la syntaxe pour l'extraction des réponses

L'évaluation de l'apport de la syntaxe lors de l'extraction de la réponse est effectuée en n'appliquant pas la technique d'unification de la variable réponse décrite à la section 2.2.1 (désactivation du module noté ② dans la figure 1).

Les résultats sont présentés dans le tableau 4. Ils montrent une hausse de 34,5 % du nombre de bonnes réponses en rang 1 en présence du module syntaxique. Enfin, le tableau 5 donne les résultats obtenus par le système en n'appliquant aucune technique d'ordre syntaxique (désactivation des modules ① et ② de la figure 1).

Si la progression obtenue est importante, il faut noter que pour un certain nombre de questions, aucune méthode "de secours", c'est-à-dire n'utilisant pas la syntaxe, n'est prévue. C'est pourquoi il est nécessaire de comparer les résultats en distinguant les types de questions ; c'est l'objectif des expérimentations décrites dans la section suivante.

| | Sans unification | Avec unification |
|---|-------------------------|-------------------------|
| Réponses correctes en rang 1 | 43.5 % | 58.5 % |
| Réponses correctes en rang 1 à 3 | 55.5 % | 70 % |
| Passages contenant la réponse en rang 1 | 55.6 % | 67.7 % |
| Passages contenant la réponse en rang 1 à 3 | 64.6 % | 75.8 % |

TAB. 4 – Apport de l'utilisation de la syntaxe pour l'extraction des réponses.

| | Sans syntaxe | Avec syntaxe |
|---|---------------------|---------------------|
| Réponses correctes en rang 1 | 36 % | 58.5 % |
| Réponses correctes en rang 1 à 3 | 51.5 % | 70 % |
| Passages contenant la réponse en rang 1 | 47 % | 67.7 % |
| Passages contenant la réponse en rang 1 à 3 | 61.1 % | 75.8 % |

TAB. 5 – Apport de l'utilisation de la syntaxe pour l'extraction des réponses.

3.4 Apport de la syntaxe selon les types de réponses

Dans le but d'évaluer de façon plus précise l'apport du module d'extraction de la réponse par unification des relations syntaxiques, nous avons isolé les questions pour lesquelles le système identifie un type d'entité nommée (EN) à rechercher. En effet, pour ces questions, en l'absence du module ②, le système va simplement rechercher une entité nommée du bon type dans le passage sélectionné. En revanche, pour les autres questions, l'unification syntaxique étant le seul moyen d'extraire une réponse candidate, le débranchement du module conduit à une absence de réponse. La progression globale constatée dépend donc fortement de la finesse de la classification en entités nommées et de l'analyse des questions correspondante.

Sur les 200 questions de la campagne CLEF 2005, 169 ont été identifiées comme attendant une entité nommée³. Par ailleurs, il est également intéressant de distinguer les questions dites factuelles des questions de définition.

Le tableau 6 résume les résultats sur ces trois classes (non mutuellement exclusives) de questions, en comparant les proportions de bonnes réponses renvoyées en première position. Il est particulièrement notable que même lorsque l'information sur l'EN attendue est présente, l'analyse syntaxique permet un gain de 14,6 %.

| | Sans unification | Avec unification |
|-----------------------------|-------------------------|-------------------------|
| Questions sur une EN | 52.7 % | 60.4 % |
| Questions factuelles | 48.3 % | 53 % |
| Questions définition | 35.3 % | 78 % |

TAB. 6 – Apport de l'utilisation de la syntaxe par types de questions (réponses en rang 1).

³Ce chiffre peut varier fortement selon des systèmes où la typologie des EN utilisées et l'analyse des questions peuvent être radicalement différentes. Ainsi, la question *Qui est Javier Solana ?* attend des EN de type "fonction", "profession", etc. car l'étiqueteur est en mesure de les détecter (Rosset *et al.*, 2007).

3.5 Conclusion

Le système FIDJI présente une architecture et un fonctionnement représentatifs des briques classiques utilisées par les systèmes de questions-réponses classiques, à ceci près qu'il fait appel en complément à des connaissances d'ordre syntaxique sur les documents interrogés et les questions posées.

Il est ainsi intéressant de procéder à une évaluation détaillée de l'apport de l'analyse syntaxique, d'une part à différents niveaux de la chaîne de questions-réponses (sélection des documents, puis extraction de la réponse), mais également selon les différents types de questions.

Nous avons présenté dans cet article une série d'expérimentations montrant que l'analyse syntaxique, combinée à des méthodes robustes et plus classiques (moteur de recherche, étiquetage en entités nommées...), pouvait conduire à une progression tout à fait significative à chaque étape du traitement.

La prochaine étape va consister, dans le cadre du projet Quaero⁴, à tester notre système sur une collection de documents provenant du Web (2 millions de pages) et ainsi évaluer ses performances sur de très grandes collections de documents de styles et de formes très différents.

Références

- BOURIGAULT D. & FABRE C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de Grammaire*, **25**.
- HARTRUMPF S., GLÖCKNER I. & LEVELING J. (2008). University of hagen at qa@clef 2008 : Efficient question answering with question decomposition and multiple answer streams. In P. FORNER, A. PEÑAS, I. ALEGRIA, C. FORASCU, N. MOREAU, P. OSENOVA, P. PROKOPIDIS, P. ROCHA, B. SACALEANU, R. SUTCLIFFE & E. T. K. SANG, Eds., *Working Notes for the CLEF 2008 Workshop*, Aarhus, Denmark.
- HATCHER E. & GOSPODNETIĆ O. (2004). *Lucene in Action*. Manning.
- KATZ B., BORCHARDT G. & FELSHIN S. (2005). Syntactic and semantic decomposition strategies for question answering from multiple resources. In *Actes de AAAI 2005 Workshop on Inference for Textual Question Answering*, Pittsburgh.
- KATZ B. & LIN J. (2003). Selectively using relations to improve precision in question answering. In *Actes de Workshop on Natural Language Processing for Question Answering, EACL*, Budapest.
- LAURENT D., SÉGUÉLA P. & NÈGRE S. (2005). Cross lingual question answering using qristal for clef 2005. In *Working Notes QA@CLEF*, Vienna.
- LAURENT D., SÉGUÉLA P. & NÈGRE S. (2006). Cross lingual question answering using qristal for clef 2006. In *Working Notes QA@CLEF*, Alicante.
- LIGOZAT A. (2007). Apport de l'analyse syntaxique des phrases dans un système de questions-réponses. *Traitement Automatique des Langues*, **46**.
- MORICEAU V., TANNIER X. & GRAU B. (2009). Utilisation de la syntaxe pour valider les réponses à des questions par plusieurs documents. In *Actes Conférence en Recherche d'Information et Applications, CORIA*, Presqu'île de Giens.

⁴<http://www.quaero.org>

ROSSET S., GALIBERT O., ADDA G. & BILINSKI E. (2007). The limsi qast systems : comparison between human and automatic rules generation for question-answering on speech transcriptions. In *ASRU*, Kyoto.

VALLIN A., GIAMPICCOLO D., AUNIMO L., 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 QA@CLEF*, Vienna.