

QUESTION GENERATION GUIDELINES

TASK DESCRIPTION: WHAT IS NEW

1. **GOAL:** Find the single paragraph that answers a question in natural language.
2. **TASK SCENARIO:** the questioner is a citizen interested in gaining insight in European legislation. The questions must be understood "according to EU legislation".
3. **DOCUMENT COLLECTION:** Subset of JRC-Acquis¹ corpus about European legislation that has parallel translations aligned at paragraph level in Bulgarian, Dutch, English, French, German, Italian, Portuguese, Romanian and Spanish. This subset is available at the ResPubliQA website: <http://celct.isti.cnr.it/ResPubliQA/Downloads>
4. **QUESTIONS:** a pool of 500 independent questions that can be answered by one single paragraph in the collection. As a difference with the previous QA tasks at CLEF, the questions are not grouped by topics, there are no NIL questions, and in the case of LISTS, all the requested items in the answer must be included in the paragraph to be returned.
5. **ANSWERS:** each question must receive one of the following two responses:
 - a) A paragraph with the candidate answer. Paragraphs to be returned are explicitly marked in the documents by <p> and their corresponding identifiers in the Acquis collection.
 - b) The string NOA to indicate that the system prefers not to answer the question.
6. **LANGUAGES INVOLVED:** Basque (EU), Bulgarian (BG), Dutch (NL), English (EN), French (FR), German (DE), Italian (IT), Portuguese (PT), Romanian (RO) and Spanish (ES).

All the monolingual² and bilingual³ combinations of questions will be activated, including Monolingual English. Basque (EU) has been included exclusively as a source languages, as there is no Basque collection available (which means that no monolingual EU-EU sub-task could be enacted).

¹ <http://wt.jrc.it/lt/Acquis/>

² Queries, target document collection, and responses are formulated in the same language.

³ Queries are formulated in a different language than the target document collection and the corresponding responses.

IMPORTANT DATES

Question Generation	from 2nd March to 27th March
Corrections	by 15th April
Translations of questions	by 4th May
Proof-reading of the question sets	8th May-18th May
Test set release	25th May
Run submissions	5th June
Assessments of runs	10th June-10th July
Results to the participants	15th July
Contributions for the Overview	by 14th August
CLEF Workshop	30 September – 02 October 2009, Corfu, Greece

Test Sets Release: May 25, 2009 (12.01 pm –noon- Central European Summer Time).

The participants will have 5 DAYS to upload their submissions, starting from the moment when the questions are downloaded.

Final Deadline for Run Submission: June 5, 2009 (11:59 pm –midnight- Central European Summer Time).

PARTNERS INVOLVED IN THE TRACK ORGANIZATION

- **ES:** Universidad Nacional de Educación a Distancia, Spain (UNED) – (general Co-ordination)
- **IT:** Center for the Evaluation of Language and Communication Technologies, Trento, Italy (CELCT) – (General Co-ordination)
- **BG:** Bulgarian Academy of Sciences, Sofia, Bulgaria (BTB)
- **EN:** University of Limerick, Ireland (LIM)
- **FR:** ELDA/ELRA, Paris, France (ELDA)
- **RO:** UAIC and RACAI, Romania.
- **EU:** University of Basque Country, Spain (UBC)

I. OUTLINE AND DATES

- **stage 1) *Creation of the Questions* (by 27th March)**

Each organizing group (CELCT, BTB, LIM, ELDA, UAIC, UBC) produces 100 questions in its own language, searches for answers in the ACQUIS parallel collection, and provides also the English translation of the questions.

N.B.: 600 candidate questions will be obtained in total, and 500 questions will be chosen out of the 600 produced.

For more details about the questions collection see section II.

- **stage 2) *Corrections* (by 15th April)**

The English translations of the 500 selected questions are checked by native speakers at CELCT and corrections from each group will be requested.

- **stage 3) *Translations of questions* (by 4th May)**

The obtained pool of questions is then translated into all the other involved languages using English as inter-lingua.

Basque, Dutch, German, and Portuguese will translate all the 500 questions into their own languages, meanwhile the groups which have already produced 100 questions in their own language will translate the remaining 400.

- **stage 4) *Proof-reading* (by 4th May)**

Each question set will be proof-read by each language group in order to check spelling and wording errors, before putting test sets online.

Like in the previous year, web-based interfaces will be provided to help in the creation and translation of the questions.

More details, together with the URL, will be provided at the beginning of the question generation phase.

N.B.: Deadlines must be considered as hard deadlines. Please try to meet them!

II. BUILDING THE QUESTION SET

As the task has been changed having a different corpus and requiring an entire paragraph as an answer, this year there is the opportunity of creating a greater variety of questions and experimenting new criteria for developing them.

Each organizing group (CELCT, BTB, LIM, ELDA, UAIC, UBC) filling in the online interface:

- ❖ produces 100 questions in its own language
- ❖ provides the English translation of the questions
- ❖ searches for the paragraph containing the answer in the respective ACQUIS parallel collection
- ❖ copies the id of the paragraph (*p_id*) in its own source language collection
- ❖ copies the id of the document from which the paragraph has been extracted
- ❖ searches for the paragraph containing the answer in the ACQUIS English parallel documents
- ❖ copies the id of the paragraph (*p_id*) in the English collection

The online interface presents the following fields:

- Question
- Translation of source questions into English
- Question type (See section III)
- Temporal restriction (YES or NO)
- Response/paragraph_string (See section IV)
- Document id and paragraph id (*doc_id* and *p_id*)
- English paragraph, (*e_p_string*)
- English paragraph id (*e_p_id*)
- Repeat with the next question up to 100.

See the section III for the description of the types of questions.

RECOMMENDATIONS

- Please pay attention at the way in which questions are expressed: try to avoid artificial questions that suggest back-formulation of statement in the original source document making questions appear somehow unnatural . Questions must be well-formed, as posed by a person that does not know the answer nor its position in the text. In principle, the word overlapping between the question and the target document should be kept at a minimum.
- Questions must be created from the <body> of the Acquis document – not from passages tagged as <note> which has no unique identifier.

III. QUESTION TYPE DESCRIPTION

Questions must NOT be grouped in series and must NOT contain anaphoric links to other entities. Each question must be guaranteed to have at least one paragraph containing the answer in the document collection (NO NIL questions should be produced).

As far as the question types are concerned, the following question categories will be considered:

- a) **Factoid**
- b) **Definition**
- c) **Reason**
- d) **Purpose**
- e) **Procedure**

- a) **Factoid questions** are fact-based questions, asking for the name of a person, a location, the extent of something, the day on which something happened, etc.

Examples:

Q: *What is the minimum angle between the generated light beams and a headlamp's reference axis?*

P: The visibility of the illuminating surface, including its visibility in areas which do not appear to be illuminated in the direction of observation considered, must be ensured within a divergent space defined by generating lines based on the perimeter of the illuminating surface and forming an angle of not less than 5° with the axis of reference of the headlamp. The origin of the angles of geometric visibility is the perimeter of the projection of the illuminating surface on a transverse plane tangent to the foremost part of the lens of the headlamp.

Q: *In how many languages is the Official Journal of the Community published?*

P: The Official Journal of the Community shall be published in the four official languages.

- b) **Definition questions** are questions such as "What/Who is X?", questions asking for the role/job/important information about someone, questions asking for the mission/full name/important information about an organization.

Examples:

Q: *What does IPP denote in the context of the environmental policies?*

P: Since then, new policy approaches on sustainable goods and services have been developed. These endeavours undertaken at all political levels have culminated in the Green Paper on Integrated Product Policy(1) (IPP). This document proposes a new strategy to strengthen and refocus product-related environmental policies and develop the market for greener products, which will also be one of the key innovative elements of the sixth environmental action programme - Environment 2010: "Our future, our choice"(2).

c) Reason questions: questions asking for the reasons/motives/motivations for something happening

Examples:

Q: *Why should the Regulation (EC) 1254 from 1999 be codified?*

P: (1) Commission Regulation (EC) No 562/2000 of 15 March 2000 laying down detailed rules for the application of Council Regulation (EC) No 1254/1999 as regards the buying-in of beef [2] has been substantially amended several times [3]. In the interests of clarity and rationality the said Regulation should be codified.

Q: *Why have the imports of live poultry from Romania been suspended?*

P: (2) Commission Decision 2005/710/EC of 13 October 2005 concerning certain protection measures in relation to highly pathogenic avian influenza in Romania [5] provides that Member States are to suspend imports of live poultry, ratites and farmed and wild feathered game and hatching eggs of those species from the whole territory of Romania and of certain products from birds from parts of that territory.

d) Purpose questions: questions asking for the aims/goals/objectives of something

Examples:

Q: *What is the Commission's aim in establishing financial forecasts each year, broken down by category of expenditure, for the three subsequent financial years?*

P: In order to place the budget of the Communities within a framework of forward planning for several years, the Commission shall, each year, after receiving the Opinion of the Budgetary Policy Committee, draw up a financial forecast for the three subsequent financial years, showing the financial implications for the Community resulting from Regulations and Decisions in force and from proposals submitted by the Commission to the Council. The forecast shall be broken down by category of expenditure.

Q: *What is the goal of the eco-label?*

P: The overall objective of the eco-label is to promote products which have the potential to reduce negative environmental impacts, as compared with the other products in the same product group, thus contributing to the efficient use of resources and a high level of environmental protection. In doing so it contributes to making consumption more sustainable, and to the policy objectives set out in the Community's sustainable development strategy (for example in the fields of climate change, resource efficiency and eco-toxicity), the sixth environmental action programme and the forthcoming White Paper on Integrated Product Policy Strategy.

e) **Procedure questions:** questions asking for a set of actions which is the official or accepted way of doing something:

Examples:

Q: *What is the procedure to import an animal to the Community from a non-member country?*

P: Animals imported from non-member countries must, with respect to the requirements for their welfare, whilst being reared, receive treatment at least equivalent to that guaranteed by this Directive to animals of Community origin. This requirement must be stated in a certificate issued by the competent authority of the non-member country concerned.

N.B.: Although this year, **LIST questions** themselves are not considered as a question category explicitly, all the above types of questions may require a list of items as an answer, provided all the requested elements are **contained in one paragraph**.

Example:

Q: *Which Republics will access to the European Union according to the Stabilisation and Association Agreement?*

P: to the Stabilisation and Association Agreement between the European Communities and their Member States, of the one part, and the Republic of Croatia, of the other part, to take account of the accession of the Czech Republic, the Republic of Estonia, the Republic of Cyprus, the Republic of Hungary, the Republic of Latvia, the Republic of Lithuania, the Republic of Malta, the Republic of Poland, the Republic of Slovenia and the Slovak Republic to the European Union.

All types of questions may contain a **temporal restriction**, i.e. a temporal specification that provides important information for the retrieval of the correct answer. Examples:

Q: *What does the approval to the Agency Rules of 5 May 1960 determine?*

P: *Approval is given to the Agency Rules of 5 May 1960 determining the manner in which demand is to be balanced against the supply of ores, source materials and special fissile materials.*

Q: *About what did the Council of 4 December 1962 decide?*

P: *COUNCIL DECISION of 4 December 1962 on the co-ordination of policies on the structure of agriculture.*

N.B.: Try to balance the number of questions per type as much as possible.

Since we aim at investigating systems' performance in answering all the above mentioned categories of questions, the test sets must be balanced, which means they should cover all the categories instead of addressing just a few of them. In summary, the indicative distribution of the questions could be **20% (20/100)** for each question category:

As mentioned above, the suggested breakdown is only indicative and each group can decide to favour a particular category/categories, and include more questions of a certain type in the question set.

IV. RESPONSES

All questions must be answered returning **one single paragraph** containing the relevant information. **Paragraphs are delimited by the mark <p>** in the collection.

- Each paragraph returned by the system is required to be an extract from a document in the parallel corpus, so the passage string must be simply cut and pasted from the corresponding document. No automatic generation of information contained in different sections of the relevant document is allowed.
- Each paragraph must contain the answer to the question. The selected paragraph must provide enough context in order to let the manual assessors decide if the answer is responsive or not.

- There is no restriction concerning the paragraph length, but the entire passage delimited by the <p> tags must be returned. Even very long paragraphs cannot be truncated.

Each paragraph returned must be associated to:

- the id of the paragraph [*p_id*] from which the passage has been cut/pasted in the source language ACQUIS collection.
- the id of the document [*doc_id*] of the ACQUIS aligned corpus from which the paragraph has been extracted;

For the sake of completeness also the corresponding English paragraph – together with its *p_id* – must be cut and pasted.

N.B.: Please make sure that also in the English collection the answer is contained in one single paragraph.