

1. Task definition

Given a set of web search results for a person name, cluster the pages according to the different people sharing the name and extract certain biographical attributes for each person (i.e., for each cluster of documents). Unlike the WePS-2 task, attributes have to be assigned to each person profile rather than to individual pages.

Groups can choose to perform only the clustering task, or both tasks together.

2. Biographical attributes

Specifically, attributes used in the WePS3 evaluation will be:

- Date of birth
- Birth place
- Other name
- Occupation
- Affiliation
- Award
- School
- Major
- Degree
- Mentor
- Nationality
- Relatives
- Phone
- FAX
- Email
- Web Site

Please refer to the [WePS-3 Attribute Extraction Task Guidelines](#) for a detailed definition of each attribute. Note two modifications with respect to WePS-2: (i) WePS2 training data had an attribute "education", which was separated into three attributes "school", "degree" and "major" in the test data. WePS-3 will use school/degree/major as independent attributes; (ii) The

annotated data in WePS-2 included "work" and "location", but these were NOT used in the WePS-2 evaluation and will not be considered in WePS-3.

2. Test Data

The test data will be composed of 300 person names and 200 web documents for each name.

As we did in WePS-2, some person names will be obtained from the following sources: US Census, Wikipedia and CS PC lists. In addition to that, we will provide names for which at least one person is an attorney, corporate executive or realtor.

The total list of names is composed of:

- 50 names from US Census.
- 50 names from Wikipedia.
- 50 names from some Computer Science Program Committee list.
- 50 names where there is at least one attorney for each name.
- 50 names where there is at least one corporate executive for each name.
- 50 names where there is at least one realtor for each name.

For each name the top 200 web search results from Yahoo! will be provided (URL, HTML pages, search snippets and ranking information).

3. Training data

We will NOT give out any training data. The evaluation data will be created in the same manner as the WePS-2 data. For the attributes, the definition of the attributes is exactly the same as the WePS-2 test data (not WePS-2 training data). Also note that in the WePS3 evaluation, participants are expected to create the attributes for each cluster, not for each document. Participants can rely on the [WePS-2 public data](#) to develop their systems.

4. Submission format

Both the clustering and attribute extraction output must be provided in the same XML file (see an example below). In this file each cluster of documents is specified by the element “entity”, which contains the list of grouped documents and the list of extracted attributes. For each attribute it's required to indicate the type of attribute (date_of_birth, occupation, etc.), the source from which it was extracted (document ranking) and the value.

```
<clustering searchString="AMANDA LENTZ"> <entity id="16" notes=""> <documents>
<doc rank="17" notes="" /> <doc rank="66" notes="" /> <doc rank="73" notes="" />
<doc rank="51" notes="from Huron" /> </documents> <attributes> <attr
type="date_of_birth" source="17" notes="">4th August 1979</attr>
<attr type="occupation" source="17" notes="">Painter</attr>
</attributes>
</entity>
[...]
```

Definition of the format, examples of usage and scripts to verify the syntactic correctness of submission files will be provided in the WePS-3 homepage (nlp.uned.es/weps) before the end of January 2010.

5. Assessments

Document Clustering

Systems are requested to make clusters as accurate as possible for the whole set of documents. However, because of the annotation load, we will evaluate only on two people per person name. In each case, at least one person will belong to one of the six above-mentioned categories.

We will make sensible manual selections in deciding which two people to select for each name, i.e. for each cluster we will select two people who have some reasonable amount of information available in the results set.

Evaluation will be done using WePS-2 metrics: Bcubed precision & recall, combined with Van Rijsbergen's F measure, and using UIR (Unanimous Improvement Ratio) to consider robustness with respect to changes in the alpha parameter which controls the weights assigned to precision and recall. See [Artiles et. al 2009](#) for details.

Attribute extraction

Attribute extraction will be done on the selected two people. Systems are required to extract values for each attribute. Participating systems will be evaluated based on the attributes they

attach to the cluster which has the best F-measure in the clustering task for the desired person. These attributes will be evaluated based on F-measure. The systems are requested to report the document ID from which they extracted each attribute value.

The attribute extraction task evaluation will be done by a pool of the system outputs, so coverage is not guaranteed on the attribute annotations.

6. Deadlines

- Release of trial data 15 February 2010
- Release of test data 7 June 2010
- Submissions due 21 June 2010
- Release of official results . 15 July 2010
- Papers due 15 August
- Workshop 23 September (CLEF 2010, Padova)