

# VISCERAL Retrieval2 Benchmark

## Guidelines for Participation v1.0

### Document History

v1.0 - 20150619 - Initial version of document

## 1. Introduction

### 1.1 Registration

The first step in participation is registration. This is done online on the following page:

<http://visceral.eu:8080/register/Registration.xhtml>

Select the “Retrieval2” benchmark on the registration screen.

During the registration process, participants will be required to sign and upload a participation agreement. Once the participant is registered and the participation agreement has been accepted by the organisers, the account will be activated. Once the account is activated, logging into the registration system will reveal the *participant dashboard*.

### 1.2 Participant Dashboard

Virtual Machines are not provided for this Benchmark, so all information about them should be ignored on the Dashboard. The main purpose of the Dashboard is to serve as a repository for documents about the Benchmark.

Documents about the Benchmark, including a list of volumes can be found by pressing


View Benchmark Files

. The evaluation software that will be used for calculating the evaluation metrics can be downloaded (see Section 6).

### 1.3 Data

Instructions for downloading the data via FTP are provided in a document found by pressing

View Benchmark Files

A list of all files in the training data set can be downloaded from the participant dashboard in the registration system by clicking on .

### 1.4 Further Information

Detailed information on the evaluation protocol and goals can be found in VISCERAL Deliverable 4.2<sup>1</sup>.

Further information on the dataset and the file formats can be found in VISCERAL Deliverable 2.2.2<sup>2</sup>.

---

<sup>1</sup> VISCERAL Deliverable 4.2:

<http://www.visceral.eu/assets/Uploads/Deliverables/VISCERAL-D4.2.pdf>

<sup>2</sup> VISCERAL Deliverable 2.2.2:

<http://www.visceral.eu/assets/Uploads/Deliverables/VISCERAL-D2.2.2.pdf>

All participants and organisers are automatically registered to the [participants-retrieval2@visceral.eu](mailto:participants-retrieval2@visceral.eu) mailing list, and can post on the mailing list. Use this list to communicate only among participants and the organisers, to ask questions, draw attention to problems or share hints and tips.

A LinkedIn group has been set-up for discussion about the Benchmark. Ask questions and make comments on this group:

<http://www.linkedin.com/groups/VISCERAL-Benchmark-Discussion-5089631>

## 2. Benchmark Organisation

### 2.1 Resources provided and their use

Participants should download the training and testing data, as well as the evaluation metric calculation software (see Section 5). Results of the tasks should only be calculated using TREC\_EVAL (see Section 5).

### 2.2 Result publication

When results obtained using the VISCERAL Retrieval2 Benchmark Resources are published, please do the following:

- Link the publication from <http://bibsonomy.org> with the tag *visceral-retrieval2*.
- Reference the following paper:  
Oscar Alfonso Jiménez-del-Toro, Antonio Foncubierto-Rodríguez, Henning Müller, Georg Langs and Allan Hanbury, Overview of the VISCERAL Retrieval Benchmark 2015, Proceedings of the Multimodal Retrieval in the Medical Domain (MRMD) Workshop, Springer LNCS 9059, 2015

## 3. VISCERAL Retrieval dataset

The VISCERALretrieval data set consists of 2311 volumes originated from three different modalities (CT,MRT1,MRT2). For a subset of these volumes we provide from the volume's radiology report extracted anatomy-pathology terms in form of csv files. The following table gives an overview of the dataset in which a participant should perform the retrieval task.

Modality	Bodyregion	Volumes	available A-P Term Files
CT	Ab	336	213
CT	Th	971	699
CT	ThAb	84	84
CT	unknown	211	211
CT	Wb	412	412
MRT1	Ab	167	114
MRT1	unknown	24	24
MRT2	Ab	68	18
MRT2	unknown	38	38
<b>TOTAL</b>		<b>2311</b>	<b>1813</b>

The volumes are available in NIFTI file format with the following naming convention:  
<volumeID>\_<modality>\_<bodyregion>.nii.gz

Anatomy - pathology term files are named accordingly as  
<volumeID>\_<modality>\_<bodyregion>.csv

An anatomy-pathology term file lists pathological terms that occur in the report of a volume together with its anatomy. Both entities are described textually and additionally with their corresponding Radlex ID (RID). Radlex is a unified language of radiology terms that can be used for standardized indexing and retrieval of radiology information resources. Detailed information about the Radlex ontology can be found at <http://rsna.org/RadLex.aspx>. A term file lists both, occurring and negated pathologies for the report. In the following we give an example to illustrate the naming convention and the file content specification.

```
VolumeID_Modality_Bodyregion.fcsv  
i.e. 123456_MRT1_Ab.csv
```

```
# The first row depicts the header of an a-p term file.  
Anatomy RID,Anatomy,Pathology RID,Pathology,Negated  
#  
# Each row entry in this file contains an occurring pathology and its anatomy and  
states if the pathology is negated.  
RID199,Ductus choledochus,RID4865,Ödem,1  
RID58,Leber,RID3874,Raumforderung,0
```

This file indicates that there is **no “Ödem” in “ductus choledochus”** whereas there occurs **“Raumforderung” in the anatomy “Leber”**.

## 4. Content-based medical image retrieval

### 4.1 Task description

In this task, we evaluate the retrieval of relevant cases based on a query case. It serves the following scenario: a user is assessing a query case in a clinical setting, e.g., a CT volume, and is searching for cases that are relevant in this assessment. The algorithm has to find cases that are relevant in a large database of cases. For each topic (query case) there is:

- the patient 3D imaging data (CT, MRI)\*
- 3D bounding box region of interest containing the radiological signs of the pathology\*
- binary mask of the main organ affected\*
- radiologic report extracted anatomy-pathology terms in the form of csv files.

\* Volumes in NIFTI file format: .nii.gz

The participants have to develop an algorithm that finds clinically-relevant (related) cases given a query case (imaging and text data), but not the final diagnosis. Relevance criterion is the relevance for differential diagnosis.

## 4.2 Evaluation

For each topic, the algorithms should generate a ranked list of search results out of the VISCERAL Retrieval dataset (containing imaging data and text data). A set of 10 test query cases (topics) is used to evaluate the result quality of the algorithms based on the ranking of cases retrieved.

The database contains both imaging data (3D in slices) and corresponding text data. There are two query scenarios:

- Image data and ROI available for the query case
- Image data, ROI and text data available for the query case.

While the first case is of immediate clinical relevance, we expect also the second case to be valuable in evaluating specific retrieval algorithms. Therefore the use of text information is optional during retrieval evaluation.

The algorithms will be evaluated in two groups corresponding to the information used for the retrieval **[image+ROI]** or **[image+ROI+text]**.

Medical experts have performed relevance assessment of the top ranked cases by each approach, to judge the quality of retrieval. The following metrics should always be calculated, as is done in the VISCERAL Retrieval Overview paper mentioned above:

- mean average precision (MAP);
- geometric mean average precision (GM-MAP);
- binary preference (bpref);
- precision after 10 cases retrieved (P10);
- precision after 30 cases retrieved (P30).

## 4.3 Retrieval test queries download

The test data (retrieval test queries) can be accessed from the same FTP service as the training data.

## 5. Result files

TREC\_EVAL should be used for calculating metrics of the result ranking. TREC\_EVAL is the standard tool used by the information retrieval community for evaluating an ad hoc retrieval run, given the results file and a standard set of judged results. The participant submissions need to be compliant with the trec\_eval format prior to submission. Result rankings that do not meet the required format are rejected. Program download and more information about TREC\_EVAL can be found at: [http://trec.nist.gov/trec\\_eval/](http://trec.nist.gov/trec_eval/).

### trec\_eval format

The format for submitting results is based on the trec\_eval program ([http://trec.nist.gov/trec\\_eval/](http://trec.nist.gov/trec_eval/)). It requires the following parameters to be arranged in

columns separated by a tab:

Ranking output	
Parameter	Explanation
<code>-topic</code>	Run number within the topic (test query): 1-10
<code>-iter</code>	Iteration number, for the VISCERAL Retrieval benchmark this number should always be 1
<code>-volumeID</code>	VolumeID from the retrieval training dataset without the extension (.nii.gz or .fcsv) and without any image path
<code>-rank</code>	Ranking result within the dataset for the corresponding query case (1-300)
<code>-simScore</code>	Similarity score assigned by the participants algorithm to perform the ranking
<code>-runID</code>	runID should be the same per run for each result ranking

<code>topic</code>	<code>iter</code>	<code>volumeID</code>	<code>rank</code>	<code>simScore</code>	<code>runID</code>
--------------------	-------------------	-----------------------	-------------------	-----------------------	--------------------

## 5.2 Specification: Ranking results

```
resultRanking_runID_ParticipantID.txt
```

```

1          1          123563_MRT1_Ab          1          0.567162          visc123
1          1          652431_CT_Wb          2          0.441542          visc123
.....
.....
.....
1          1          764393_CT_undefined          1000          0.035022          visc123
2          1          833887_MRT2_Ab          1          0.94753           visc123
.....
.....
.....
10         1          4553266_CT_Wb          995          0.014921          visc123

```

.....: Rest of the results in the ranking. Notice there are no spaces between rows.

The score should be in decreasing order (i.e. the image at the top of the list should have a higher score than images at the bottom of the list).