TECHNISCHE
UNIVERSITÄT
DARMSTADT

Tristan Miller

Presented at:
School of Data Analysis and Artificial Intelligence
National Research University – Higher School of Economics
25 May 2017

Overview

TECHNISCHE
UNIVERSITÄT
DARMSTADT

What are shared tasks?

What is CodaLab Competitions?

Organizing CodaLab Competitions

Student competitions

Caveats

What are shared tasks?

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

▶ define a data processing **task** (classification, segmentation, ranking, etc.)

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

► define a data processing **task** (classification, segmentation, ranking, etc.)
► define **evaluation metrics** to measure performance

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ▶ define a data processing **task** (classification, segmentation, ranking, etc.)
- ▶ define **evaluation metrics** to measure performance
- ▶ produce **test data** and the **gold standard** answers

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ▶ define a data processing **task** (classification, segmentation, ranking, etc.)
- ▶ define **evaluation metrics** to measure performance
- ▶ produce **test data** and the **gold standard** answers

- ▶ solicit participants to write **algorithms** to process the test data

Introduction

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ▶ define a data processing **task** (classification, segmentation, ranking, etc.)
- ▶ define **evaluation metrics** to measure performance
- ▶ produce **test data** and the **gold standard** answers

- ▶ solicit participants to write **algorithms** to process the test data

- ▶ applying the evaluation metrics, **score** the algorithms' output against the gold standard

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ▶ define a data processing **task** (classification, segmentation, ranking, etc.)
- ▶ define **evaluation metrics** to measure performance
- ▶ produce **test data** and the **gold standard** answers
    - ▶ produce **trial data** for demonstration purposes
    - ▶ produce or provide **ancillary resources** (knowledge bases, etc.)
    - ▶ produce **training data** for use by supervised algorithms
- ▶ solicit participants to write **algorithms** to process the test data

- ▶ applying the evaluation metrics, **score** the algorithms' output against the gold standard

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ► define a data processing **task** (classification, segmentation, ranking, etc.)
- ► define **evaluation metrics** to measure performance
- ► produce **test data** and the **gold standard** answers
  - ► produce **trial data** for demonstration purposes
  - ► produce or provide **ancillary resources** (knowledge bases, etc.)
  - ► produce **training data** for use by supervised algorithms
- ► solicit participants to write **algorithms** to process the test data
  - ► implement various **baseline algorithms**
- ► applying the evaluation metrics, **score** the algorithms' output against the gold standard

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

In a **shared task** / **challenge** / **competition** / **evaluation campaign** / **evaluation exercise**, the organizers

- ▶ define a data processing **task** (classification, segmentation, ranking, etc.)
- ▶ define **evaluation metrics** to measure performance
- ▶ produce **test data** and the **gold standard** answers
    - ▶ produce **trial data** for demonstration purposes
    - ▶ produce or provide **ancillary resources** (knowledge bases, etc.)
    - ▶ produce **training data** for use by supervised algorithms
- ▶ solicit participants to write **algorithms** to process the test data
    - ▶ implement various **baseline algorithms**
- ▶ applying the evaluation metrics, **score** the algorithms' output against the gold standard
    - ▶ compare and analyze the participants' and baseline algorithms

Shared tasks: pros and cons

Pros:

- ► stimulate methodological research on unsolved problems
- ► provide standardized data sets, resources, and evaluation metrics
- ► facilitate reproducibility of research results
- ► centralize publication and discussion of research results

Shared tasks: pros and cons

Pros:

- ▶ stimulate methodological research on unsolved problems
- ▶ provide standardized data sets, resources, and evaluation metrics
- ▶ facilitate reproducibility of research results
- ▶ centralize publication and discussion of research results

Cons:

- ▶ everything must be planned in advance
- ▶ large organizational overhead (data distribution, publicity, communication with participants, etc.)
- ▶ encourages "teaching to the test"

What is CodaLab Competitions?

- ▶ Web-based platform for running online data-based competitions
- ▶ Developed by Microsoft, Stanford University, and others
- ▶ Free hosted implementation: `https://competitions.codalab.org/`
- ▶ Free software (Apache License 2.0):
  `https://github.com/codalab/codalab-competitions/`

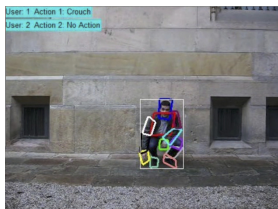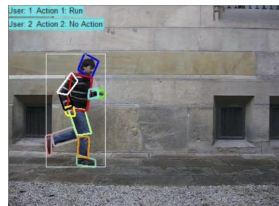# Microsoft COCO Image Captioning Challenge
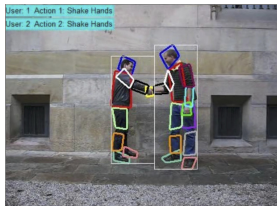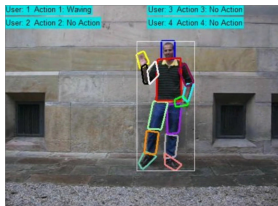

The man at bat readies to swing at the pitch while the umpire looks on.

A large bus sitting next to a very tall building.

CodaLab features
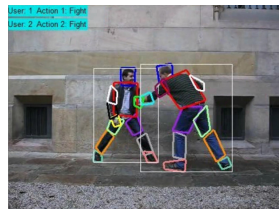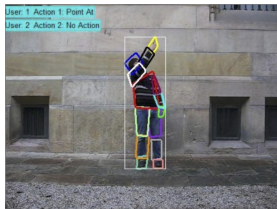
- ▶ Hosts public task website
- ▶ Hosts private gold-standard data and scoring software
- ▶ Manages participant registration
- ▶ Enforces submission deadlines
- ▶ Runs scoring software
- ▶ Tabulates, stores, and publishes results
- ▶ Handles communication between organizers/participants (e-mail, forums)
- ▶ Provides some publicity

CodaLab

Accelerating reproducible computational research.

Popular Competitions

	LiTS - Liver Tumor Segmentation Challenge	Dec 01, 2016 **439** participants
	Microsoft COCO Image Captioning Challenge	Mar 15, 2015 **333** participants
	CIKM AnalytiCup 2017: Lazada Product Title	May 01, 2017 192 participants

Featured Competitions

	ChaLearn LAP Real Versus Fake Expressed Emotion	Apr 20, 2017 22 participants
	MICCAI Multimodal Brain Tumor	Aug 30, 2013 94 participants
	The Large Scale Movie Description Challenge (LSMDC)	Aug 31, 2016 11 participants

Competitions

language   [Search]

The Story Cloze Test
Organized by ROCNLP

A challenge for evaluating a system's natural language and story understanding.

Aug 30, 2016-*No end date*

30 participants

The Large Scale Movie Description Challenge (LSMDC) 2016 : Movie Multiple-Choice Test
Organized by atousa

The goal of the challenge is to evaluate different visual-language models performance to annotate videos based on natural sentences for ...

Aug 25, 2016-*No end date*

18 participants

The Large Scale Movie Description Challenge (LSMDC) 2016 : Movie Retrieval
Organized by atousa

The goal of the challenge is to evaluate different visual-language models performance to retrieve videos based on natural sentence queries ...

Aug 25, 2016-*No end date*

17 participants

Competition for SemEval-2017/Task-11
Organized by julsal

End-User Development using Natural Language

Aug 01, 2016-Feb 04, 2017

Competition

SemEval-2017 Task 7, Subtask 3

Organized by Logological · Current server time: May 7, 2017, 2:11 p.m. UTC

▶ **Current**   Next

Trial   Test (Homographic)

Dec. 5, 2016, midnight UTC   Jan. 23, 2017, midnight UTC

Learn the Details   Phases   Participate   Results   Forums ⦿

- Overview
- Evaluation
- Terms and Conditions

Overview

This is the CodaLab Competition for Subtask 3 of SemEval-2017 Task 7: Detection and Interpretation of English Puns.

Background

A pun is a form of wordplay in which one signifier (*e.g.*, a word or phrase) suggests two or more meanings by exploiting polysemy, or phonological similarity to another signifier, for an intended humorous or rhetorical effect. For example, the first of the following two punning jokes exploits contrasting meanings of the word "interest", while the second exploits the sound similarity between the surface form "propane" and the latent target "profane":

> I used to be a banker but I lost interest.
>
> When the church bought gas for their annual barbecue, proceeds went from the sacred to the propane.

Competition

SemEval-2017 Task 7, Subtask 3

Organized by Logological · Current server time: May 7, 2017, 2:11 p.m. UTC

▶ Current	Next
Trial	Test (Homographic)
Dec. 5, 2016, midnight UTC	Jan. 23, 2017, midnight UTC

Learn the Details   Phases   Participate   Results   Forums ➦

- Overview
- **Evaluation**
- Terms and Conditions

Evaluation criteria

The evaluation for this subtask will be carried out in two simultaneous phases, one for the homographic data set and one for the heterographic data set. Systems may participate in either or both phases.

Systems participating in a given phase may provide single guess for any or all of the contexts in the data set.

The results for each phase must be submitted in a delimited text file named `answer.txt`. Each line of the text file consists of three fields separated by horizontal whitespace (a single tab or space character). The first field is the ID of a pun word from the data set. The second field is a semicolon-delimited list of WordNet 3.1 sense keys that match one meaning of the pun. The third field is a semicolon-delimited list of WordNet 3.1 sense keys that match the other meaning of the pun. Sample data and results files are available in the trial data.

To submit the results, place `answer.txt` in a ZIP file (in the top-level directory), and then upload it to CodaLab according to the instructions at Participating in a Competition.

Competition

SemEval-2017 Task 7, Subtask 3

Organized by Logological · Current server time: May 7, 2017, 2:11 p.m. UTC

▶ **Current**   Next

Trial   Test (Homographic)

Dec. 5, 2016, midnight UTC   Jan. 23, 2017, midnight UTC

Learn the Details   **Phases**   Participate   Results   Forums ⏏

Trial

Start: Dec. 5, 2016, midnight

Test (Homographic)

Start: Jan. 23, 2017, midnight

Test (Heterographic)

Start: Jan. 23, 2017, midnight

Competition Ends

Jan. 30, 2017, 11:59 p.m.

Competition

Admin features

Edit   Participants   Submissions

SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

▶ Current	Next
Trial	Test (Homographic)
Dec. 5, 2016, midnight UTC	Jan. 23, 2017, midnight UTC

Learn the Details   Phases   **Participate**   Results   Forums ➕

Get Data

Submit / View Results

Data sets

Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Download

- Trial and test data

Note that, due to the difficulty in amassing a large number of pun examples per word or per sense, there is **no training data** for this task.

Learn the Details   Phases   **Participate**   Results   Forums 🔀

Get Data

Submit / View Results

Trial   Test (Homographic)   Test (Heterographic)

Phase description
None

Max submissions per day: 999

Max submissions total: 999

Click the Submit button to upload a new submission.

Optionally add more information about this submission

Submit

Here are your submissions to date (✔ indicates submission on leaderboard):

#	FILENAME	SUBMISSION DATE	STATUS	✔	
1	allwrong.zip	12/19/2016 13:06:30	Finished		➕
2	alsocorrect.zip	12/19/2016 13:06:32	Finished		➕
3	badfields.zip	12/19/2016 13:06:35	Failed		➕
4	badid.zip	12/19/2016 13:06:37	Failed		➕
5	correct.zip	12/19/2016 13:06:40	Finished		➕
6	dup.zip	12/19/2016 13:06:43	Failed		➕
7	halfcoverage.zip	12/19/2016 13:06:45	Finished		➕

SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

▶ Current   Next

Trial   Test (Homographic)

Dec. 5, 2016, midnight UTC   Jan. 23, 2017, midnight UTC

Learn the Details   Phases   Participate   **Results**   Forums ➕

Trial   Test (Homographic)   Test (Heterographic)

Phase description
None

Max submissions per day: 999

Max submissions total: 2

📊 Download CSV   Download all submissions on leaderboard

#	User	Team Name	F1 ▲	recall ▲	precision ▲	coverage ▲
1	doogyb		0.0774 (4)	0.0770 (4)	0.0778 (4)	0.9900 (1)
2	kevang	BuzzSaw	0.1544 (2)	0.1525 (1)	0.1563 (2)	0.9761 (2)
3	lhurtado	ELiRF-UPV	0.0996 (3)	0.0978 (3)	0.1014 (3)	0.9646 (3)
4	evrog	UTMN	0.0452 (5)	0.0424 (5)	0.0484 (5)	0.8760 (4)
5	tpederse		0.1557 (1)	0.1448 (2)	0.1683 (1)	0.8606 (5)

Organizing CodaLab Competitions

Competitions (My CodaLab)

Competitions | Competitions I'm In | **Competitions I'm Running** | My Datasets

Create Competition

SemEval-2017 Task 7, Subtask 3
Organized by Logological

Interpretation of English Puns

Dec 05, 2016-Jan 30, 2017

15 participants

Edit | Unpublish | Participants | Submissions | Leaderboard

SemEval-2017 Task 7, Subtask 2
Organized by Logological

Location of English Puns

Dec 05, 2016-Jan 22, 2017

16 participants

Edit | Unpublish | Participants | Submissions | Leaderboard

SemEval-2017 Task 7, Subtask 1
Organized by Logological

Detection of English Puns

Dec 05, 2016-Jan 15, 2017

16 participants

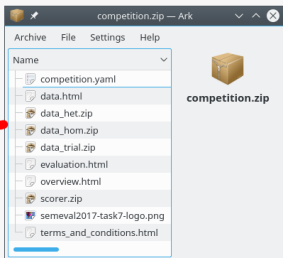Edit | Unpublish | Participants | Submissions | Leaderboard

Create Competition

To create a competition, upload a bundle with the definition of the new competition. ❓

Upload definition...

Create Competition

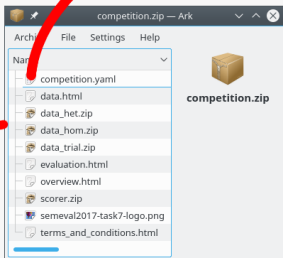To create a competition, upload a bundle with the definition of the new competition. ❓

Upload definition...

competition.zip — Ark

Archive   File   Settings   Help

Name ▾

- competition.yaml
- data.html
- data_het.zip
- data_hom.zip
- data_trial.zip
- evaluation.html
- overview.html
- scorer.zip
- semeval2017-task7-logo.png
- terms_and_conditions.html

competition.zip

Competition bundle file structure

Competitions are defined by a zipped file archive (a **bundle**) containing:

- ▶ a logo image
- ▶ HTML files for the competition website
- ▶ scoring software (ZIP archive)
- ▶ gold-standard ("reference") data (ZIP archive)
- ▶ `competition.yaml`

`competition.yaml` **settings**

- ▶ Competition title, description, and logo
- ▶ Whether registration is required
- ▶ Filenames for standard and option web pages
- ▶ Configuration of competition phases
- ▶ Format of the leaderboard

`competition.yaml`: **basic settings**

```yaml
title: SemEval-2017 Task 7, Subtask 3
description: Interpretation of English Puns
image: semeval2017-task7-logo.png
has_registration: True
allow_teams: True
end_date: 2017-01-30
html:
    overview: overview.html
    evaluation: evaluation.html
    terms: terms_and_conditions.html
    data: data.html
⋮
⋮
```

Competition

Admin features

Edit   Participants   Submissions

SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

▶ Current	Next
Trial	Test (Homographic)
Dec. 5, 2016, midnight UTC	Jan. 23, 2017, midnight UTC

Learn the Details   Phases   **Participate**   Results   Forums ⊕

Get Data

Submit / View Results

Data sets

Data sets for this subtask are described in detail on the SemEval-2017 Task 7 website.

Download

- Trial and test data

Note that, due to the difficulty in amassing a large number of pun examples per word or per sense, there is **no training data** for this task.

HTML files

```html
<p>Data sets for this subtask are described in detail on the
    <a href="http://alt.qcri.org/semeval2017/task7/index.php
   ?id=data-and-resources">SemEval-2017 Task 7 website</a>.<
   /p>

<h4>Download</h4>
<ul>
 <li><a href="data/uploads/semeval2017_pun_task.tar.xz">
    Trial data</a></li>
 <li>Test data will not be released until the evaluation
    begins</li>
</ul>

<p>Note that, due to the difficulty in amassing a large
   number of pun examples per word or per sense, there is <
   strong>no training data</strong> for this task.</p>
```

Competition phases

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ▶ **Phases** break your competition into optional subtasks
- ▶ Each phase has its own:
  - ▶ title
  - ▶ start date
  - ▶ scoring program
  - ▶ data
  - ▶ submission limit
- ▶ Phases can be run in parallel, staggered, or (with some difficulty) in sequence
- ▶ It's common to have a "trial" phase for sandbox testing:
  - ▶ earlier start date
  - ▶ toy data
  - ▶ unlimited submissions

`competition.yaml`: **defining phases**

```
phases:
    1:
          phasenumber: 1
          label: "Trial"
          start_date: 2016-12-05
          max_submissions: 999
          scoring_program: scorer.zip
          reference_data: data_trial.zip
          leaderboard_management_mode: hide_results
          color: white
    2:
          phasenumber: 2
          label: "Test␣(Homographic)"
          start_date: 2017-01-23
          ⋮
```

Competition

SemEval-2017 Task 7, Subtask 3

Organized by Logological · Current server time: May 7, 2017, 2:11 p.m. UTC

▶ Current	Next
Trial	Test (Homographic)
Dec. 5, 2016, midnight UTC	Jan. 23, 2017, midnight UTC

Learn the Details   **Phases**   Participate   Results   Forums ⏻

Trial

Start: Dec. 5, 2016, midnight

Test (Homographic)

Start: Jan. 23, 2017, midnight

Test (Heterographic)

Start: Jan. 23, 2017, midnight

Competition Ends

Jan. 30, 2017, 11:59 p.m.

Leaderboards

- A **leaderboard** is a dynamically updated results table
- Each phase has 0 or more leaderboards, public or hidden
- The columns are the metrics output by your scoring software
- The rows are the participants' submissions
- You define the column labels and numeric format
- You can also rank the metrics by priority

`competition.yaml`: **defining the leaderboards**

```
leaderboard:
    leaderboards:
        RESULTS: &RESULTS
            label: Results
            rank: 1
    columns:
        coverage:
            leaderboard: *RESULTS
            label: coverage
            rank: 4
            numeric_format: 4
        precision:
            leaderboard: *RESULTS
            label: precision
            ⋮
            ⋮
```

SemEval-2017 Task 7, Subtask 3

Organized by Logological - Current server time: May 7, 2017, 2:14 p.m. UTC

▶ Current	Next
Trial	Test (Homographic)
Dec. 5, 2016, midnight UTC	Jan. 23, 2017, midnight UTC

Learn the Details   Phases   Participate   **Results**   Forums ➕

Trial   Test (Homographic)   Test (Heterographic)

Phase description
None

Max submissions per day: 999

Max submissions total: 2

📊 Download CSV   Download all submissions on leaderboard

Results

#	User	Team Name	F1 ▲	recall ▲	precision ▲	coverage ▲
1	doogyb		0.0774 (4)	0.0770 (4)	0.0778 (4)	0.9900 (1)
2	kevang	BuzzSaw	0.1544 (2)	0.1525 (1)	0.1563 (2)	0.9761 (2)
3	lhurtado	ELiRF-UPV	0.0996 (3)	0.0978 (3)	0.1014 (3)	0.9646 (3)
4	evrog	UTMN	0.0452 (5)	0.0424 (5)	0.0484 (5)	0.8760 (4)
5	tpederse		0.1557 (1)	0.1448 (2)	0.1683 (1)	0.8606 (5)

- ▶ You provide a ZIP file containing arbitrary gold-standard **reference data**
- ▶ You provide another ZIP file containing
  - ▶ an executable **scorer**
  - ▶ a **metadata** file that describes how to run the scorer, using the following variables:

$program	scorer directory
$input/ref	reference data directory
$input/res	submission data directory
$output	output directory

```
description: Scoring program for Subtask 3 of SemEval
    -2017 Task 7 (pun interpretation)
command: java -classpath $program de.tudarmstadt.ukp.
    semeval2017.task7.scorer.PunScorer -i $input/ref/
    truth.txt $input/res/answer.txt $output/scores.txt
```

Scoring programs and reference data

TECHNISCHE
UNIVERSITÄT
DARMSTADT
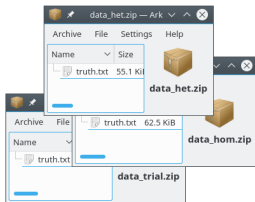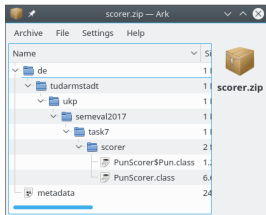
- Scorers can be written in any language (Python, Java, Perl, etc.)
- The scorer must produce a key–value file $output/scores.txt
- Each key corresponds to a leaderboard column key:

```
coverage: 1.000
precision: 0.825
recall: 0.775
f1: 0.799
```

- stderr is captured and reported to the submitter
- stderr and stdout are captured and stored for the competition organizer

data.html — KWrite

```html
<h3>Data sets</h3>

<p>
    Data sets for this subtask are described in detail on
    the <a
       href="http://alt.qcri.org/semeval2017/task7/
       index.php?id=data-and-resources">SemEval-2017
       Task 7 website</a>.
</p>

<h4>Download</h4>

<ul>
    <li><a
       href="http://alt.qcri.org/semeval2017/task7/data/
       uploads/semeval2017_pun_task.tar.xz">Trial
       data</a></li>
    <li>Test data will not be released until the
    evaluation begins</li>
</ul>
<p>
    Note that, due to the difficulty in amassing a large
    number of pun
    examples per word or per sense, there is <strong>no
    training
    data</strong> for this task.
</p>
```
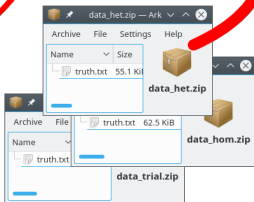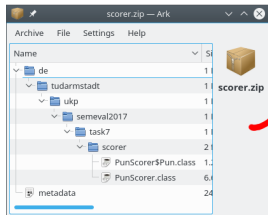
competition.yaml — KWrite

```yaml
title: SemEval-2017 Task 7, Subtask 3
description: Interpretation of English Puns
image: semeval2017-task7-logo.png
has_registration: True
allow_teams: True
end_date: 2017-01-30
html:
    overview: overview.html
    evaluation: evaluation.html
    terms: terms_and_conditions.html
    data: data.html
phases:
    1:
        phasenumber: 1
        label: "Trial"
        start_date: 2016-12-05
        max_submissions: 999
        scoring_program: scorer.zip
        reference_data: data_trial.zip
        leaderboard_management_mode: hide_results
        color: white
    2:
        phasenumber: 2
        label: "Test (Homographic)"
        start_date: 2017-01-23
        max_submissions: 2
        scoring_program: scorer.zip
        reference_data: data_hom.zip
        color: blue
    3:
        phasenumber: 3
        label: "Test (Heterographic)"
        start_date: 2017-01-23
        max_submissions: 2
        scoring_program: scorer.zip
        reference_data: data_het.zip
        color: purple
leaderboard:
    leaderboards:
        RESULTS: &RESULTS
            label: Results
            rank: 1
    columns:
        coverage:
            leaderboard: *RESULTS
            label: coverage
            rank: 4
            numeric_format: 4
        precision:
            leaderboard: *RESULTS
            label: precision
            rank: 3
```

scorer.zip — Ark

Archive   File   Settings   Help

Name	Si
∨ ■ de	1
∨ ■ tudarmstadt	1
∨ ■ ukp	1
∨ ■ semeval2017	1
∨ ■ task7	1
∨ ■ scorer	2
PunScorer$Pun.class	1.
PunScorer.class	6.
■ metadata	24

data_het.zip — Ark

Archive   File   Settings   Help

Name	Size
▼ truth.txt	55.1 KiB

data_hom.zip

Archive   File

| ▼ truth.txt | 62.5 KiB |

Name	∨

data_trial.zip

Archive   File

Name	∨
▼ truth.txt	

data.html — KWrite

```
<h3>Data sets</h3>

<p>
    Data sets for this subtask are described in detail on
    the <a
        href="http://alt.qcri.org/semeval2017/task7/
        index.php?id=data-and-resources">SemEval-2017
        Task 7 website</a>.
</p>

<h4>Download</h4>
<ul>
    <li><a
        href="http://alt.qcri.org/semeval2017/task7/data/
        uploads/semeval2017_pun_task.tar.xz">Trial
        data</a></li>
    <li>Test data will not be released until the
    evaluation begins</li>
</ul>
<p>
    Note that, due to the difficulty in amassing a large
    number of pun
    examples per word or per sense, there is <strong>no
    training
    data</strong> for this task.
</p>
```
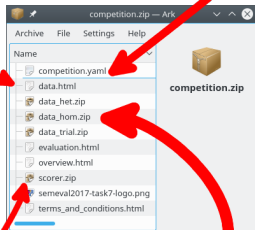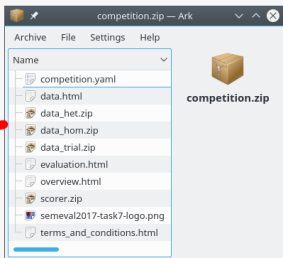
competition.yaml — KWrite

```
title: SemEval-2017 Task 7, Subtask 3
description: Interpretation of English Puns
image: semeval2017-task7-logo.png
has_registration: True
allow_teams: True
end_date: 2017-01-30
html:
    overview: overview.html
    evaluation: evaluation.html
    terms: terms_and_conditions.html
    data: data.html
phases:
    1:
        phasenumber: 1
        label: "Trial"
        start_date: 2016-12-05
        max_submissions: 999
        scoring_program: scorer.zip
        reference_data: data_trial.zip
        leaderboard_management_mode: hide_results
        color: white
    2:
        phasenumber: 2
        label: "Test (Homographic)"
        start_date: 2017-01-23
        max_submissions: 2
        scoring_program: scorer.zip
        reference_data: data_hom.zip
        color: blue
    3:
        phasenumber: 3
        label: "Test (Heterographic)"
        start_date: 2017-01-23
        max_submissions: 2
        scoring_program: scorer.zip
        reference_data: data_het.zip
        color: purple
leaderboard:
    leaderboards:
        RESULTS: &RESULTS
            label: Results
            rank: 1
    columns:
        coverage:
            leaderboard: *RESULTS
            label: coverage
            rank: 4
            numeric_format: 4
        precision:
            leaderboard: *RESULTS
            label: precision
            rank: 3
```

competition.zip — Ark

Archive File Settings Help

Name
- competition.yaml
- data.html
- data_het.zip
- data_hom.zip
- data_trial.zip
- evaluation.html
- overview.html
- scorer.zip
- semeval2017-task7-logo.png
- terms_and_conditions.html

competition.zip

scorer.zip — Ark

Archive File Settings Help

Name
- de
  - tudarmstadt
    - ukp
      - semeval2017
        - task7
          - scorer
            - PunScorer$Pun.class 1..
            - PunScorer.class 6..
  - metadata 24

scorer.zip

data_het.zip — Ark

Archive File Settings Help

Name	Size
truth.txt	55.1 Ki

data_het.zip

data_hom.zip

Archive File

Name	
truth.txt	62.5 KiB

data_hom.zip

data_trial.zip

Name
- truth.txt

data_trial.zip

Upon uploading a competition

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ▶ You will (usually) be warned if there is a problem
- ▶ You optionally "publish" your competition:

  Unpublished: accessible only if you know the URL (default)
    Published: listed on CodaLab Competitions home page

- ▶ Competitions can be edited via the web interface to:
  - ▶ Publish/unpublish
  - ▶ Change competition settings
  - ▶ Edit web pages (via rich text editor)
  - ▶ Add or remove reference data and scoring programs
- ▶ Editing the original YAML is no longer possible!

Competitions (My CodaLab)

Competitions   Competitions I'm In   **Competitions I'm Running**   My Datasets

Create Competition

SemEval-2017 Task 7, Subtask 3
Organized by Logological

Interpretation of English Puns

Dec 05, 2016-Jan 30, 2017

15 participants

| Edit | Unpublish | Participants | Submissions | Leaderboard |

SemEval-2017 Task 7, Subtask 2
Organized by Logological

Location of English Puns

Dec 05, 2016-Jan 22, 2017

16 participants

| Edit | Unpublish | Participants | Submissions | Leaderboard |

SemEval-2017 Task 7, Subtask 1
Organized by Logological

Detection of English Puns

Dec 05, 2016-Jan 15, 2017

16 participants

| Edit | Unpublish | Participants | Submissions | Leaderboard |

Edit Competition

Download original YAML file   ❓ Help

General Information

Title:
SemEval-2017 Task 7, Subtask 3

Description:
Interpretation of English Puns

☐ Disallow leaderboard modifying

☐ Force submission to leaderboard

Logo:
Currently: logos/semeval2017-task7-logo_33.png ☐ Clear
Change:
Browse...   No file selected.

☑ Registration Required

End Date (UTC):
2017-01-30 23:59:59

☑ Publicly Available

☐ Enable medical image viewer

☐ Enable detailed results

Competitions (My Datasets)

Competitions   Competitions I'm In   Competitions I'm Running   My Datasets

Create Dataset

NAME	KEY	DESCRIPTION	TYPE	ACTIONS	MULTI-DELETE? (UNSAFE!)
scorer_1_15705	66628265-3f45-4a9f-bb08-ca21ed02b33b	None	Scoring Program	Download Delete	☐
data_trial_1_15705	ced44e66-4c5c-4a40-86ad-e35bc6385d4b	None	Reference Data	Download Delete	☐
scorer_1_15706	fc3a8853-0490-47c8-b3f4-9cb82c7e40b4	None	Scoring Program	Download Delete	☐
data_trial_1_15706	d81e2ebf-e875-431d-8c16-281670513146	None	Reference Data	Download Delete	☐
scorer_1_15707	96adea9b-0957-4e00-84ff-8cddb01ec5fd	None	Scoring Program	Download Delete	☐
data_trial_1_15707	6159a63c-e180-40b8-9ec7-6594a0482b97	None	Reference Data	Download Delete	☐
scoring_program_1_15847	ae8efe2a-d979-444a-bdf3-7316b9d6ba9d	None	Scoring Program	Download Delete	☐
dev_data_1_15847	d8277bc8-3660-4329-ad47-f09fe610c5bb	None	Reference Data	Download Delete	☐
test_data_2_15847	a666e75d-869d-4dde-bfe1-4fb727c85dde	None	Reference Data	Download Delete	☐
SemEval 2017 Task 7 Subtask 1 Homographic	bf68f74d-02c1-4682-b1ce-33894c1b5082		Reference Data	Download	☐
SemEval 2017 Task 7	eb0c1b99-cf6f-4171-			Download Delete	☐

Viewing and downloading submissions

TECHNISCHE
UNIVERSITÄT
DARMSTADT

The "Submissions" page allows you to:

- ▶ see a list of all successful and failed submissions
- ▶ download the original system output
- ▶ download a CSV of all system scores
- ▶ re-run the scorer on one or all submissions
- ▶ delete a submission
- ▶ (un)hide a submission from the leaderboard

Submissions

Phase

Go Back to Competition Homepage

Test (Homographic) - Jan. 23, 2017, midnight UTC

View all submission's metadata

Select phase description

None

Max submissions per day: 999

Max submissions total: 2

Submissions

[x] Download CSV   Re-run all submissions in this phase

#	SUBMITTED	SUBMITTED BY	SUBMISSION ID	FILENAME	STATUS	LEADERBOARD	RESULTS					
1	Jan. 25, 2017, 5:41 p.m.	kevang	385636	WSDse_split.zip	Finished	True	0.9761	+	DEL	HIDE	FAILED	RE-RUN
2	Jan. 27, 2017, 3:34 a.m.	evrog	385821	homs_3.zip	Finished	True	0.8760	+	DEL	HIDE	FAILED	RE-RUN
3	Jan. 27, 2017, 7 p.m.	lhurtado	385895	answer.txt_task7.3.zip	Failed	False	None	+	DEL	SHOW	FAILED	RE-RUN
4	Jan. 27, 2017, 7:10 p.m.	lhurtado	385897	answer.txt_task7.3.zip	Finished	True	0.9646	+	DEL	HIDE	FAILED	RE-RUN
5	Jan. 28, 2017, 4:02 a.m.	evrog	385931	homs_3_extended.zip	Failed	False	None	+	DEL	SHOW	FAILED	RE-RUN
6	Jan. 28, 2017, 4:20 a.m.	evrog	385932	homs_3_extended.zip	Finished	False	None	+	DEL	SHOW	FAILED	RE-RUN

Student competitions

université
PARIS-SACLAY

Projet M2 & L2 AIC
2016/2017

Mini Challenges 2017

View Presentation

ParisSaclay Center for Data Science

CHA LEARN

CodaLab

Microsoft

Challenges (M2 students)	Abstract	Task	Solution (Video L2 students)
Blue	**Activity of molecules against HIV** The problem is to relate molecular structure to activity to screen new compounds before actually testing them with High Throughput Screening (HTS) in vitro experiments. HTS is a method for massive scientific experimentation used in drug discovery, linking the fields of biology and chemistry. This method remains very costly process despite many recent technological advances in the field of biotechnology. This is why applying machine learning methods would be of great benefit for the pharmaceutical industry to reduce the number of compounds that need to be tested.	The Objective of is to predict which compounds are active against the AIDS HIV infection. The dataset has two classes : active or inactive (Binary Classification). The variables represent properties of the molecule inferred from its structure. Note: this project is running on the LRI server. In case of problem, a **previous version** on the main Codalab instance is available.	Marine Cobalt
Cyan	**Lothlorien** This challenge aims at addressing the issue of resources access (website, drug purchase, violent movie, etc.) based on the age of a person. Indeed a lot of violent content is accessible on the internet and 45 % of children under 12 are not monitored by parental control. For this sake, we rely on the person's	A computer vision challenge is proposed for undergraduate students in which the challenger must predict the class of a person (major or minor) based on a picture of his/her face. Note: the main Codalab instance of this challenge has been tested. Note: this project is running on the LRI server. In case of problem,	Cerulean Turquoise

Student competitions

- ▶ skill development for student participants/organizers
    - ▶ experimentation, problem formulation/solving
    - ▶ writing/fulfilling technical specifications
    - ▶ collaboration with peers
    - ▶ presentation of research results

Student competitions

- ▶ skill development for student participants/organizers
    - ▶ experimentation, problem formulation/solving
    - ▶ writing/fulfilling technical specifications
    - ▶ collaboration with peers
    - ▶ presentation of research results
- ▶ benefits for student participants
    - ▶ unrestricted choice of tools/methods
    - ▶ instant feedback to students
    - ▶ freedom to experiment

Student competitions

- ► skill development for student participants/organizers
  - ► experimentation, problem formulation/solving
  - ► writing/fulfilling technical specifications
  - ► collaboration with peers
  - ► presentation of research results
- ► benefits for student participants
  - ► unrestricted choice of tools/methods
  - ► instant feedback to students
  - ► freedom to experiment
- ► benefits for teachers
  - ► automatic enforcement of submission deadlines
  - ► instant tabulation of scores
  - ► consistent packaging of submissions

Caveats

Caveats

TECHNISCHE
UNIVERSITÄT
DARMSTADT

CodaLab Competitions is . . .

- ▶ underdocumented
- ▶ unpredictable
- ▶ unintuitive
- ▶ unstable

Caveats

TECHNISCHE
UNIVERSITÄT
DARMSTADT

CodaLab Competitions is . . .

- ▶ underdocumented
- ▶ unpredictable
- ▶ unintuitive
- ▶ unstable

But it's . . .

- ▶ free (as in beer)
- ▶ free (as in speech)
- ▶ popular
- ▶ under active development
- ▶ probably a time-saver, in the long run

Thank you!

Questions?