

Class 04: SemEval shared tasks

October 2, 2017

What is SemEval ?

- International competition for semantic analysis systems
- 20-years old
- Started because evaluating semantics was hard and systems weren't really comparable

History

| | Year | Location | Tasks |
|--------------|-------------|-----------------|--------------|
| Senseval-1 | 1998 | Sussex | – |
| Senseval-2 | 2001 | Toulouse | – |
| Senseval-3 | 2004 | Barcelona | – |
| SemEval-2007 | 2007 | Prague | 19 |
| SemEval-2010 | 2010 | Uppsala | 18 |
| SemEval-2012 | 2012 | Montreal | 8 |
| SemEval-2013 | 2013 | Atlanta | 14 |
| SemEval-2014 | 2014 | Dublin | 10 |
| SemEval-2015 | 2015 | Denver | 18 |
| SemEval-2016 | 2016 | San Diego | 14 |
| SemEval-2017 | 2017 | Vancouver | 12 |
| SemEval-2018 | 2018 | TBA | 12 |

This year

- Twelve tasks in five main areas
 - Affect and creative language in tweets
 - Coreference
 - Information extraction
 - Lexical semantics
 - Reading comprehension and reasoning
- SemEval-2018 will be the 12th workshop on semantic evaluation and will be located at [TBD].
- Languages represented:
 - English (12), Spanish (3), Italian (1), Arabic (1)

Timeline

| | |
|--------------|--|
| 25 September | Data ready, evaluation script and baseline system available. |
| 08 January | Evaluation period starts |
| 29 January | Evaluation period ends |
| 26 February | System description paper deadline |
| 2 April | Paper notifications |
| 16 April | Camera-ready submissions due |

Approximately 13 weeks to get a system up and running.

Vyshka participation

Why participate in a shared task ?

- Great learning experience
- All participants get to send a system description paper
 - Co-located with a high-level conference (NAACL, ACL, COLING, ...)
- Looks good on CV — both “participated” and “was best in x”
 - Especially if you’re planning on applying for PhD positions

What next ?

- Describe (most of) this year’s tasks
- See if any of them look interesting
- Check out some of the baseline systems

Affect and creative language in tweets

Task 1: Affect in Tweets

Previous datasets / tasks have focussed on simple binary classification of emotion. This task looks at how much of each emotion is felt.

For example, given a tweet:

- detect the emotional intensity of {anger, fear, joy, sadness} from 0..1



| | |
|-----|---------|
| 0.9 | Joy |
| 0.0 | Fear |
| 0.0 | Sadness |
| 0.1 | Anger |

Languages: English, Arabic, Spanish

Tweet ID, Tweet text, emotion, intensity

| | | | | |
|-------|---|-------|-------|-------|
| 11615 | I'm much to full of resentment | anger | 0.688 | |
| 11570 | Blood is boiling | anger | 0.729 | |
| 11571 | why aggravate me on purpose? | anger | 0.604 | |
| 11572 | @DonnyMurray I'm just no offended by stuff. | anger | | 0.479 |
| 11574 | @Matt_Lat fuming ain't the word | anger | 0.646 | |
| ... | | | | |

Task 2: Multilingual emoji prediction

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-----|----|---|---|---|---|---|---|---|---|----|----|----|----|----|----|-----|----|----|----|----|
| USA | ❤️ | 😍 | 😂 | 💕 | 🔥 | 😊 | 😎 | ✨ | 💙 | 😘 | 📷 | 🇺🇸 | ☀️ | 💜 | 😏 | 100 | 😁 | 🎄 | 📷 | 😜 |
| ESP | ❤️ | 😍 | 😂 | 💕 | 😊 | 😘 | 💪 | 😏 | 👉 | 🇪🇸 | 😎 | 💙 | 💜 | 😏 | 💕 | ✨ | 🎵 | 💕 | 😁 | 👉 |

Task, given a tweet,

- predict which emoji the tweet had beforehand.
- single emoji per tweet.

 **Josep Capdevila#DUI**
@josepcapdevila [Segueix](#) ▾

El País se supera cada día. Ahora el Major de los Mossos se llama Josep Lluís Drapaire
😂😂😂

Traduït del espanyol

EL PAÍS CATALUNYA

En el seu haver-hi, els Mossos expliquen també amb la comunicació. Malgrat la confusió de la primera jornada, la informació sobre els fets va quedar en mans no d'un polític sinó d'un tècnic, el major Josep Lluís Drapaire. El comissari cap ha transmès, coincideixen els experts, un missatge de tranquil·litat. I ha aportat, amb rigor, les dades necessàries per informar a la població, reservant aquells detalls necessaris per esclarir els fets.



 **Josep Capdevila#DUI**
@josepcapdevila [Segueix](#) ▾

El País se supera cada día. Ahora el Major de los Mossos se llama Josep Lluís Drapaire

Traduït del espanyol

EL PAÍS CATALUNYA

En el seu haver-hi, els Mossos expliquen també amb la comunicació. Malgrat la confusió de la primera jornada, la informació sobre els fets va quedar en mans no d'un polític sinó d'un tècnic, el major Josep Lluís Drapaire. El comissari cap ha transmès, coincideixen els experts, un missatge de tranquil·litat. I ha aportat, amb rigor, les dades necessàries per informar a la població, reservant aquells detalls necessaris per esclarir els fets.

Languages: US English, Spanish

Only available after registering

<https://github.com/fvancesco/Semeval2018-Task2-Emoji-Detection/>

Training data:

```
0    A little throwback with my favourite person @ Water Wall
2    Who never... @ A Galaxy Far Far Away
9    Birthday Kisses @ Madison, Wisconsin
```

Labels:

```
0    ♥    _red_heart_
1    😊    _smiling_face_with_hearteyes_
2    😊    _face_with_tears_of_joy_
...
9    😘    _face_blowing_a_kiss_
...
```

Task 3: Irony detection in English tweets

Irony: A trope or figurative language use whose actual meaning differs from what is literally enunciated.

Two tasks:

- Binary classification ironic/not ironic
- Four-way classification:
 - 0) Non-ironic: ...And it's payday. THIS IS A GOOD FRIDAY
 - 1) Polarity contrast: I love waking up with migraines #not :'(
 - 2) Situational: Event technology session is having Internet problems. #irony
 - 3) Other: @someuser Yeah keeping cricket clean, that's what he wants #Sarcasm

Languages: English

Task A: (Binary)

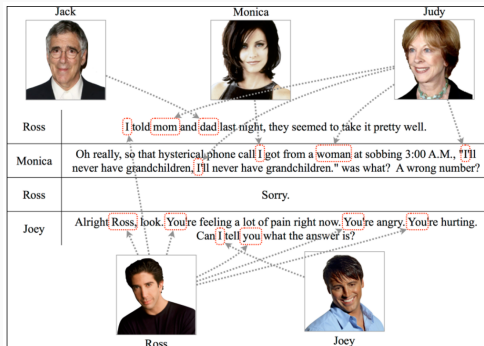
| | | |
|------|---|--|
| 1907 | 0 | It's #cute to be awake right now |
| 2415 | 1 | Wishing my relationship was that real |
| 1454 | 1 | Good thing y days not ruined.... |
| 2493 | 0 | Not this room please I don't want this |
| 2725 | 0 | @ajb_16 shouldn't you's be in school?? |
| 1294 | 0 | I'm so ready to get drunk tonight |
| 2480 | 1 | Dead supportive family I've got. |

Task B: (4-way)

| | | |
|------|---|---|
| 1907 | 0 | It's #cute to be awake right now |
| 2993 | 2 | Spilled milk onto my boob. Oh the . |
| 3527 | 2 | I cant even watch anime in japan... |
| 3504 | 3 | On a small diet http://t.co/UsLZ0gYmEx |
| 2198 | 3 | @torrentfreak We should monitor them... |
| 2480 | 1 | Dead supportive family I've got. |

Coreference

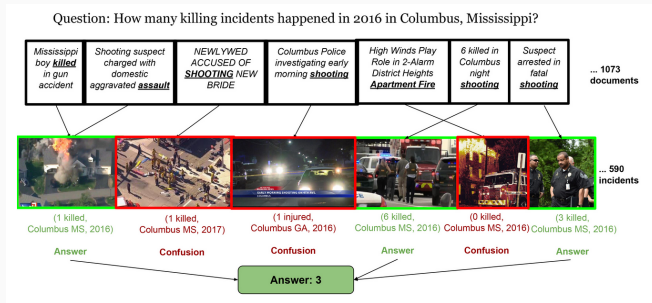
Task 4: Character identification on multiparty dialogues



| | | | | | | | | | Speaker | Entity ID |
|---|---|------|-----|-------------|------|---|---|---------------|---------|-----------|
| 0 | 0 | He | PRP | (TOP(S(NP* | he | - | - | Monica_Geller | * | (284) |
| 0 | 1 | 's | VBZ | (VP* | be | - | - | Monica_Geller | * | - |
| 0 | 2 | just | RB | (ADVP* | just | - | - | Monica_Geller | * | - |
| 0 | 3 | some | DT | (NP(NP* | some | - | - | Monica_Geller | * | - |
| 0 | 4 | guy | NN | *) | guy | - | - | Monica_Geller | * | (284) |
| 0 | 5 | I | PRP | (SBAR(S(NP* | I | - | - | Monica_Geller | * | (248) |
| 0 | 6 | work | VBP | (VP* | work | - | - | Monica_Geller | * | - |
| 0 | 7 | with | IN | (PP*)))))) | with | - | - | Monica_Geller | * | - |
| 0 | 8 | ! | . | *) | ! | - | - | Monica_Geller | * | - |

Task 5: Counting events and participants in the long tail

Given a question, find news articles that correspond to the question.



Difficulty:

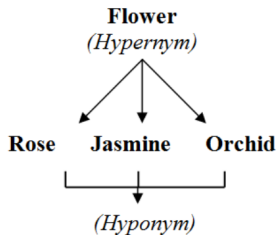
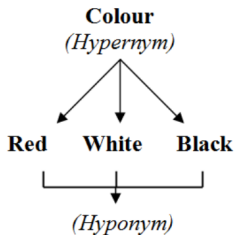
- For each correct article, there are 674 incorrect articles with overlapping context

Languages: English

Only available after registering

Lexical semantics

Task 9: Hypernym discovery



| | Term | Hypernym(s) | Source |
|-------------------|-------------|-----------------------------------|---------------------------|
| English (general) | dog | canine, mammal, animal | WordNet |
| Spanish (general) | guacamole | salsa para mojar, alimento, salsa | Wikidata |
| Italian (general) | Nina Simone | musicista, pianista, persona | MultiWiBi |

Languages: English, Spanish, Italian

Input:

pollution Concept

Output:

dirtyiness

dirtying

environmental condition

impurity

sanitary condition

uncleanness

Task 10: Capturing discriminative attributes

- Current models are great at detecting similarity
- But do not predict semantic differences
- If you only know what something is like but can't tell the difference it is a problem for your model



The task:

- Given a triple (cappuccino, americano, milk), return
 - 1 is a semantic difference
 - 0 is not a semantic difference

Languages: English

Only available after registering

Reading comprehension and reasoning

Task 11: Machine comprehension using commonsense knowledge

My garden was looking a little empty, so I decided I would plant something. I went out and bought tree seeds. I found a spot that looked like it would get enough sunshine. There, I dug a hole for the seeds. Once that was done, I took my watering can and watered the seeds.

A. Why was the tree planted?

- to get enough sunshine
- the garden looks empty
- they were forced
- the soil was moist

B. What was used to dig the hole?

- the wind
- a tablespoon
- a trowel [?]
- their bare hands

Languages: English

Task 11: Machine comprehension using commonsense knowledge

My garden was looking a little empty, so I decided I would plant something. I went out and bought tree seeds. I found a spot that looked like it would get enough sunshine. There, I dug a hole for the seeds. Once that was done, I took my watering can and watered the seeds.

A. Why was the tree planted?

- to get enough sunshine
- the garden looks empty
- they were forced
- the soil was moist

B. What was used to dig the hole?

- the wind
- a tablespoon
- a trowel [?]
- their bare hands

Languages: English

Task 11: Machine comprehension using commonsense knowledge

My garden was looking a little empty, so I decided I would plant something. I went out and bought tree seeds. I found a spot that looked like it would get enough sunshine. There, I dug a hole for the seeds. Once that was done, I took my watering can and watered the seeds.

A. Why was the tree planted?

- to get enough sunshine
- the garden looks empty
- they were forced
- the soil was moist

B. What was used to dig the hole?

- the wind
- a tablespoon
- a trowel [?]
- their bare hands

Languages: English

Only available after registering

Task 12: Argument reasoning comprehension task

Given an argument, with a reason and a claim, the goal is to choose the correct implicit reasoning from two options.

Topic: Should voting be mandatory?

Additional info: Or are there already too many people casting ballots?

Argument: It works for Australia. And since [...] voting should be mandatory.

- (a) Australians are nothing like Americans
- (b) Australians are like Americans

Languages: US English

<https://github.com/habernal/semEval2018-task12>

TSV file:

| | |
|--------------|---|
| warrant0 | Australians are nothing like Americans |
| warrant1 | Australians are like Americans |
| correctLabel | 1 |
| reason | It works for Australia. |
| claim | Voting should be mandatory |
| debateTitle | Should Voting Be Mandatory? |
| debateInfo | Or are there already too many people casting ballots? |

Practical

<http://alt.qcri.org/semeval2018/index.php?id=tasks>

- Choose a task
- Read the description
- Look at the provided data
- Run the baseline system (Irony detection, argument reasoning)
- Write your own simple baseline (Everything else)
- Run the evaluation script on the baseline
- Think of ways to improve the baseline