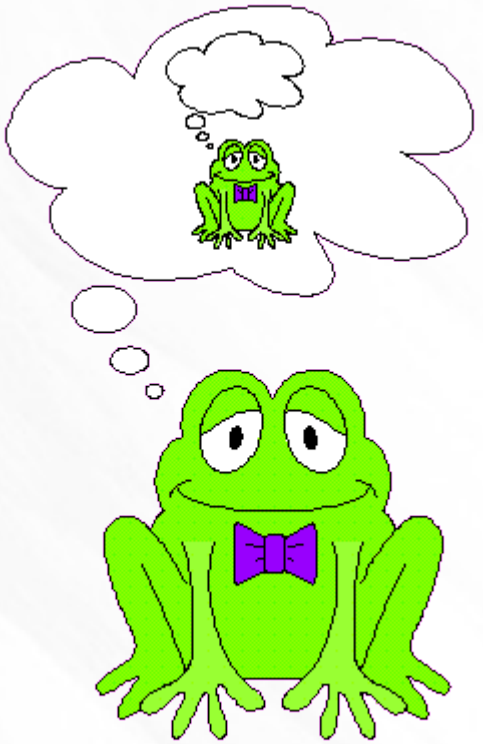
Meta-cognition evaluation

Maruseac Mihai
Neață Sofia
https://github.com/mihaimaruseac/nlp

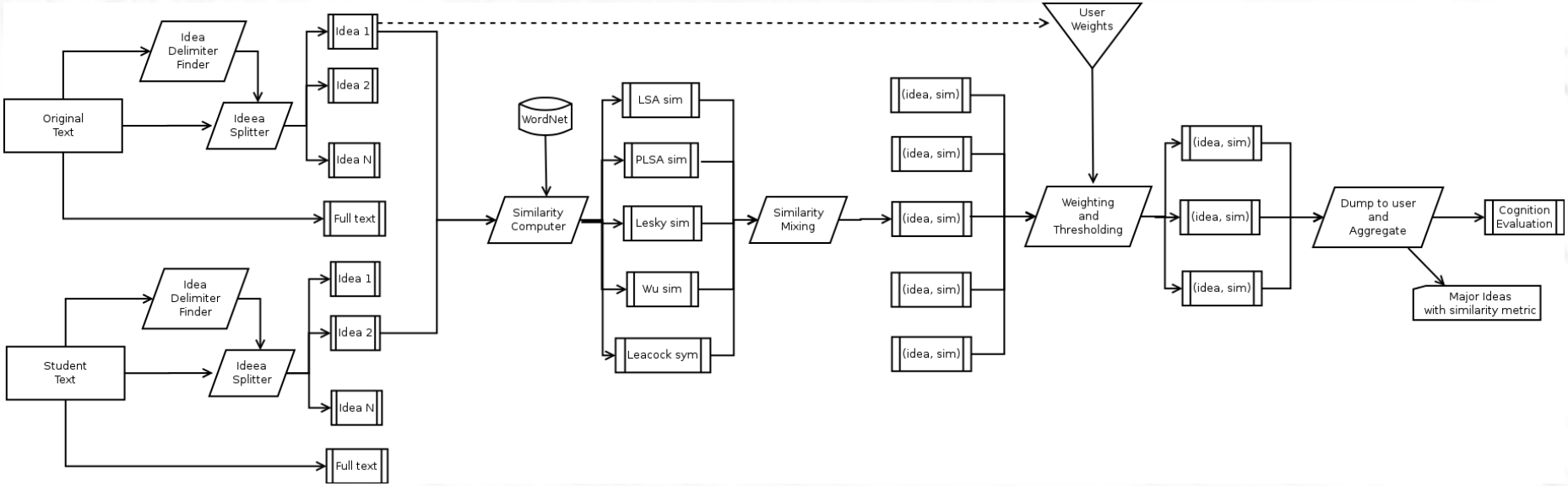
Metacognition

- Thinking about knowing
- Learning about thinking
- Control of learning



Overview

$$MC\ O\ S = presentUser\ \$\ filter\left(TF\ w_{ideeas}\right)\$\ map\ SC\ \$\ MI\left(getIdeas\ O\right)\left(getIdeas\ S\right)$$



Beta Stage

- Pre-alpha
- POS tagging
- One similarity measure
- Multiple problems

MontyLingua

```
(NX Tristan/NNP Thorn/NNP NX) (/( (NX Charlie/NNP Cox/NNP NX) )/) (VX is/VBZ VX) (NX the/DT son/NN NX) of/IN (NX Dunstan/NNP Thorn/NNP NX)
and/CC (NX a/DT captive/JJ princess/NN NX) (VX called/VBN VX) (NX Una/NNP NX) ./.
(NX The/DT couple/NN NX) (VX met/VBD only/RB VX) (NX one/CD night/NN NX) and/CC (VX fell/VBD VX) in/IN (NX love/NN NX) instantly/RB ./.
Unfortunately/RB ,/, when/WRB (NX Tristan/NNP NX) (VX was/VBD born/VBN VX) ,/, (NX she/PRP NX) (VX was/VBD not/RB allowed/VBN to/TO keep/VB
 VX) (NX him/PRP NX) and/CC (VX instead/RB sent/VBD VX) (NX him/PRP NX) to/TO (VX live/VB VX) with/IN (NX his/PRP$ father/NN NX) ./.
(NX She/PRP NX) (VX put/VBP VX) in/IN (NX his/PRP$ basket/NN NX) a/DT "/" (NX Babylon/NNP candle/NN NX) "/" and/CC (NX a/DT letter/NN NX) t
o/TO (NX Tristan/NNP NX) (VX explaining/VBG VX) (NX everything/NN NX) ./.


   SENTENCE #1 DIGEST:

         adj_phrases: []
  adj_phrases_tagged: []
           modifiers: ['captive']
    modifiers_tagged: ['captive/JJ']
        noun_phrases: ['Tristan Thorn', 'Charlie Cox', 'son', 'Dunstan Thorn', 'captive princess', 'Una']
 noun_phrases_tagged: ['Tristan/NNP Thorn/NNP', 'Charlie/NNP Cox/NNP', 'son/NN', 'Dunstan/NNP Thorn/NNP', 'captive/JJ princess/NN', 'Una/NN
P']
parameterized_predicates: [[['be', []], ['', []], ['son', ['determiner=the']], ['of Dunstan Thorn', ['prep=of']]], [['call', ['past_tense']
], ['captive princess', ['determiner=a']], ['Una', []]]]
        prep_phrases: ['of Dunstan Thorn']
 prep_phrases_tagged: ['of/IN Dunstan/NNP Thorn/NNP']
 verb_arg_structures: [['is/VBZ', '', ['son/NN', 'of/IN Dunstan/NNP Thorn/NNP']], ['called/VBN', 'a/DT captive/JJ princess/NN', ['Una/NNP']
]]
verb_arg_structures_concise: ['("be" "" "son" "of Dunstan Thorn")', '("call" "captive princess" "Una")']
        verb_phrases: ['is', 'called']
 verb_phrases_tagged: ['is/VBZ', 'called/VBN']


   SENTENCE #2 DIGEST:

         adj_phrases: []
  adj_phrases_tagged: []
           modifiers: ['only', 'instantly']
```

Language

- French in original data set
- MontyLingua supports only English

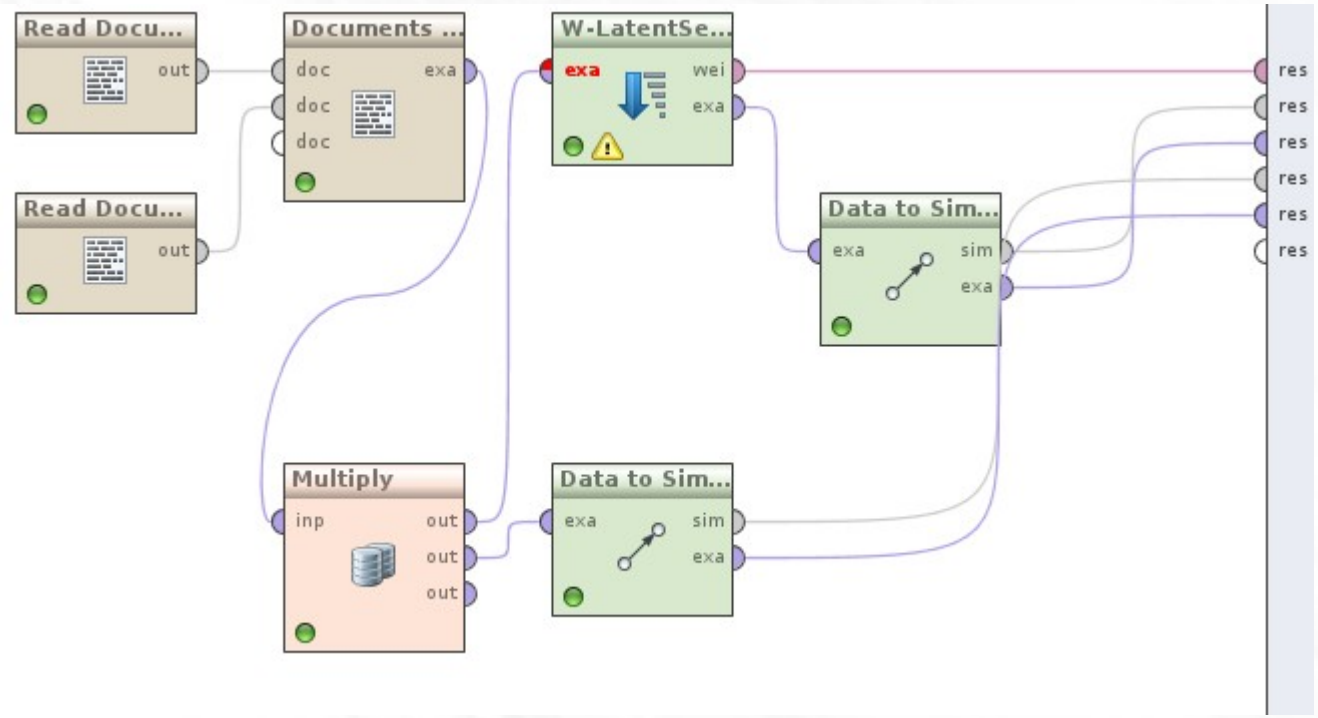
- Fallback to IMDB data
- Switch to other NLP toolchain

OpenNLP vs MontiLingua vs GATE

- MontiLingua
- Python, a collection of libraries
- Use what you need
- No training required
- Enriched with common sense knowledge
- Less vulnerable to NLP errors

- Only English???

Rapid Miner


Read Docu...
out
Read Docu...
out
Documents ...
doc
doc
doc
exa
W-LatentSe...
exa
wei
exa
Data to Sim...
exa
sim
exa
Multiply
inp
out
out
out
Data to Sim...
exa
sim
exa
res
res
res
res
res
res

RapidMiner

- All components are premade
- No tutorial, hard to understand what each of them does
- How to convert a document to a wordlist?

RapidMiner vs WEKA

- RapidMiner
- More text analysis tools
- Command line interface – usable from our application only when needed
- Multilayered-data-view concept mapping closely to our pipeline idea

- Still a good decision (after understanding how it works)

Results

- We have some results
- Distance: 200028.57
- Something must be wrong
- Need tweaking and debugging

Thanks