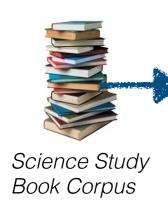
# Crowdsourcing Multiple Choice Science Questions

# Johannes Welbl, Nelson F. Liu, Matt Gardner

University College London, University of Washington, Allen Institute for Artificial Intelligence

j.welbl@cs.ucl.ac.uk, nfliu@cs.washington.edu, mattg@allenai.org

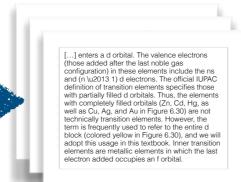
# **Stage 1: Corpus to Paragraph**



### **Document Filter**

**Want**: easy-to-understand, simple, declarative facts.

- 1. length constraints
- 2. no personal pronouns
- 3. no questions/exclamations
- 4. .



Paragraph Collection



**Stage 2: Crowdsourced Question Creation** 

Q: Elements have orbitals that are filled with what?

A: electrons

Question-Answer collection

# **Stage 3: Multiple Choice Distractor Proposal**

#### **QA-pair**

Q: Elements have orbitals that are filled with what? A: electrons

#### **Distractor Candidate**

"newspaper"

## **Distractor Classification Model**



 $P(c \text{ good distractor} \mid Q, A)$ 

Random Forest

Bad

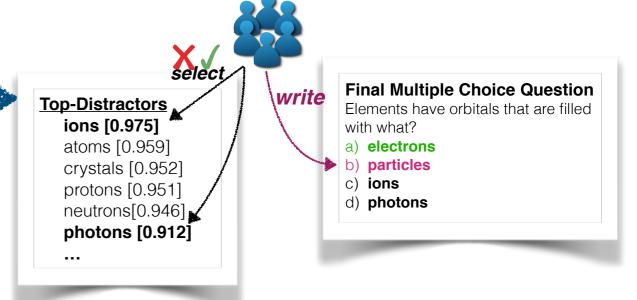
#### Features:

syntactic, distributional semantics, Knowledge Base, hypernymy, ...

#### **Desirable Distractors: Characteristics**

- syntactically consistent (e.g. plural noun)
- relevant / on-topic (here: atom models)
- type constraints (e.g. chem. element)

# **Stage 4: Question & Distractor Validation**



# Outcome: SciQ

13.7K crowdsourced multiple choice science exam questions

http://data.allenai.org/sciq/#sciq-dataset