

INTRODUCTION

As students read expository text, comprehension is improved by pausing to answer questions that reinforce the material. This project showcases a computer program that automatically generates questions from the text itself.

Applications for this technology include:

- Intelligent tutoring systems
- Self-study apps
- Questions for tutors and teachers
- Quiz and test writing assistants

SAMPLE SOURCE TEXT

The atom is a basic unit of matter that consists of a dense central nucleus surrounded by a cloud of negatively charged electrons. The atomic nucleus contains a mix of positively charged protons and electrically neutral neutrons. The electrons of an atom are bound to the nucleus by the electromagnetic force. Likewise, a group of atoms can remain bound to each other by chemical bonds based on the same force, forming a molecule.

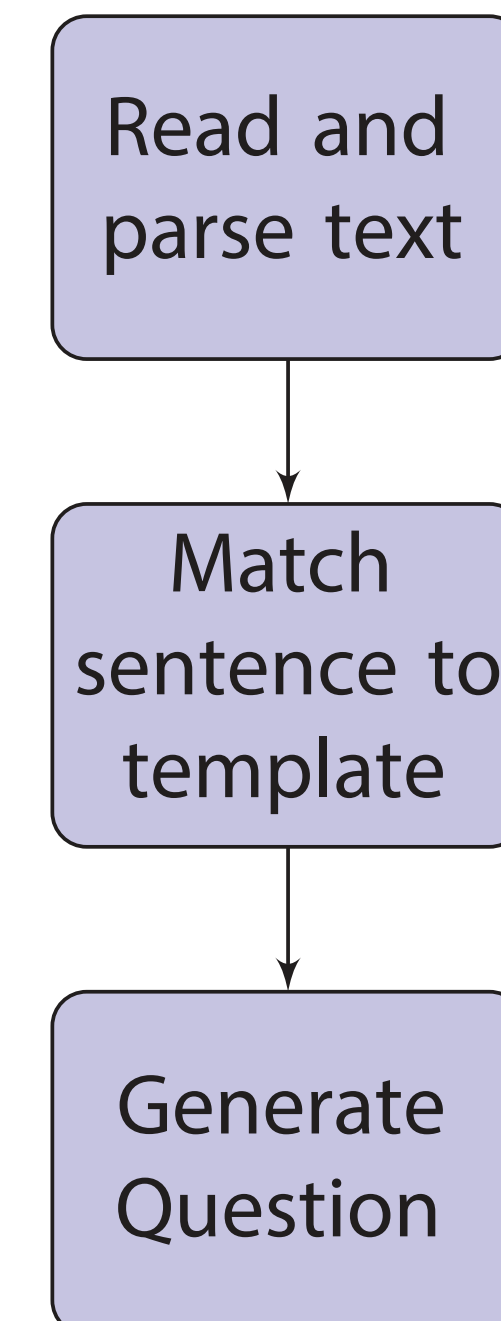
An atom containing an equal number of protons and electrons is electrically neutral, otherwise it is positively or negatively charged and is known as an ion. An atom is classified according to the number of protons and neutrons in its nucleus. The number of protons determines the chemical element, and the number of neutrons determines the isotope of the element.

IMPACT

Roediger and Pyc [5] describe multiple studies which show that students who are more frequently asked questions about the material they have read retain significantly more than those who are not. Mostow et al. [4] demonstrated that reading comprehension can be boosted with questions that are generated automatically.

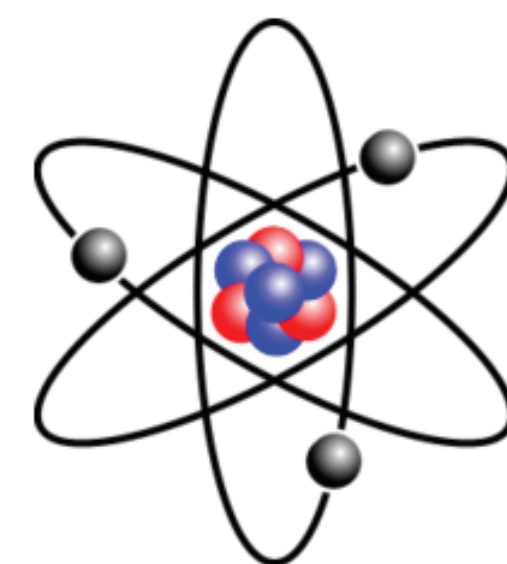
HOW DOES IT WORK ?

The system inputs plain text which is processed by parsing software. Next, the parsed sentences are compared to templates to find matches. Currently the system has > 50 templates which were manually created. Future work involves automatically generating templates from a corpus.



SAMPLE GENERATED QUESTIONS

- Describe an atom.
- What does the atomic nucleus contain?
- How would you explain the role of the electromagnetic force as described in this passage?
- Discuss the role or impact of chemical bonds.
- Explain what a group of atoms can form.
- How would you describe an atom containing an equal number of protons and electrons?
- What does the number of protons determine?



REFERENCES

- [1] Atom diagram and sample source text from Wikipedia: <http://en.wikipedia.org/wiki/Atom>
- [2] Heilman, M., Smith, N. Question generation via overgeneration transformations and ranking. Carnegie-Mellon University, No. CMU-LTI-09-013.
- [3] Mazidi, K., Nielsen, R. Pedagogical Evaluation of Automatically Generated Questions. (under review)
- [4] Mostow, J., et al. Using automated questions to assess reading comprehension, vocabulary, and effects of tutorial interventions. *Technology Instruction Cognition and Learning* 2 (2004): 97-134.
- [5] Roediger, H., Pyc, M. Inexpensive techniques to improve education. *J. Applied Research in Memory and Cognition* 1.2 (2012) 242-248.

EXAMPLE

Given the sample sentence:

The atomic nucleus contains a mix of positively charged protons and negatively charged neutrons.

First, the sentence is read and parsed as shown below. The predicate is identified, and its arguments are labelled with their semantic roles. In this sentence we have semantic role labels Agent and Patient.

the atomic nucleus	contains	a mix of positively charged protons and electrically neutral neutrons
Agent	verb	Patient

Next, the system takes this parsed sentence and looks for matching templates. The template shown below matches and so a question is generated.

Finally, template fields are filled in from the source text to form a question and answer. The verb and argument Agent are used to create the question. The argument Patient is the answer to the question.

Template:	What	do Agent	verb	?
Question:	What	does the atomic nucleus	contain	?

RESULTS

By carefully designing the templates, quality questions can be generated to assist students in understanding and remembering what they have read. In a recent evaluation [3], 500 automatically generated questions from the system were compared against 500 questions from a similar system created by a team from CMU [2]. The table below shows the evaluation results. The numbers represent the percentage of questions in each category that annotators found to be grammatical and clear (linguistic evaluation) and most helpful for remembering and understanding the text (pedagogical evaluation). Future work involves creating a machine learning ranker to select a few high-quality questions.

	Linguistic Evaluation		Pedagogical Evaluation	
	UNT System	CMU System	UNT System	CMU System
Science Text	72	54	45	32
Social Studies Text	74	55	34	27

CONTACT INFORMATION

Email KarenMazidi@my.unt.edu
Lab <http://hilt.cse.unt.edu/>
Director Rodney.Nielsen@unt.edu