U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

DEI DEPARTAMENTO DE
ENGENHARIA INFORMÁTICA

LIACC

Exploring NLP and Information Extraction to jointly address Question Generation and Answering

P. Azevedo, B. Leite, H. L. Cardoso, D. C. Silva, L. P. Reis
Artificial Intelligence and Computer Science Lab
University of Porto, Portugal

AIAI 2020 - 16th International Conference on
Artificial Intelligence Applications and Innovations
5 – 7 June, 2020

Agenda

- Introduction
- Problem
- Goals
- Literature Review
- System Overview
- Approach
- Evaluation
- Conclusions

Introduction

- **Question Generation (QG)** aims to **generate questions** from a text passage;
- **Question Answering (QA)** is concerned with building systems that automatically **answer questions**;
- **QG** and **QA** have been subjects of an intensive study in recent years and much progress has been made in both areas;
- By **combining these two fields** it will be possible to understand **how they could improve each other**.

Problem

- When combining QG and QA the **main focus** is on how QG can be used to improve QA;
- **Lack of automatic mechanisms** in order **to evaluate** the results of QG;
- **Lack of automatic mechanisms** in order to **effortlessly generate** datasets for QA;
- **Lack of approaches** that combine QG and QA in order to perceive how these tasks can improve each other.

Our goals

- Apply **AI** and **Natural Language Processing techniques (NLP)** to **generate questions** from English texts and then **answer** those same questions;
- Use QA to get a **perception of the ambiguity** from the generated questions;
- Use QG to draw conclusions about **QA robustness**;
- Understand the **potential** of a system that **combines** question generation with question answering.

Literature review for Question Generation (QG)

- **Template-based** by using pre-defined templates of the questions;
  - [Awad et al., 2014], [Le et al., 2015]
- **Syntactic Analysis** by manipulating the syntactic structure of the sentence;
  - [Majumder et al., 2015], [Danon et al., 2017]
- **Semantic Analysis** which focus on semantic parse (using Semantic Role Labeling);
  - [Araki et al., 2016], [Flor et al., 2018]
- **Dependency Analysis** which connects words in a sentence using their functional relations;
  - [Mazidi et al., 2016]
- **Machine Learning** techniques using Neural Networks (*Seq2Seq* approaches).
  - [Chen et al., 2018], [Lu et al., 2019]

Literature review for Question Answering (QA)

- **Question Analysis**;
  - Recognize query and target result types [Pakray et. al. 2011] [High 2012]
  - Phrase level dependency graph [Xu et. al. 2014]

- **Passage Retrieval**;
  - Deal as query [Unger et. al. 2012] [Hakimov et. al. 2013]
  - Knowledge Graphs [Usbeck et. al. 2015] [Shekarpour et. al. 2015]

- **Answer Extraction**
  - N-grams [Ittycheriah et. al. 2006] [Echihabi et. al. 2008]
  - Hot terms [Le et. al. 2016]

Literature review using both QG and QA

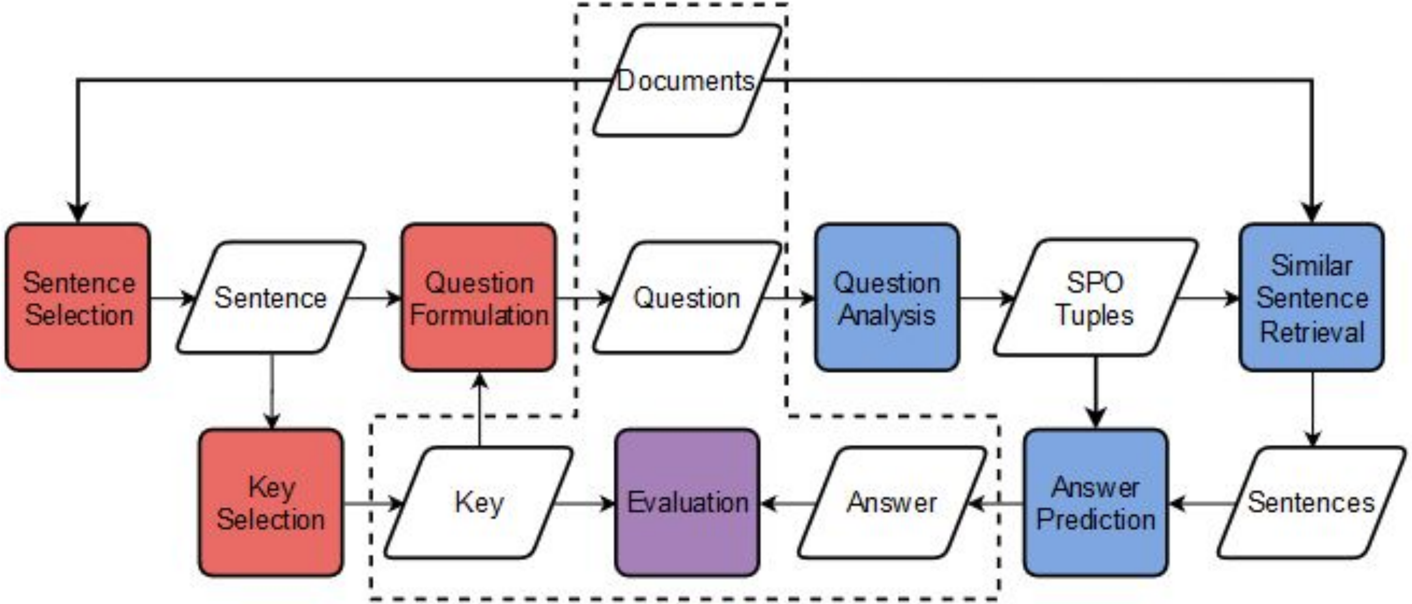
- (**QG helps QA**) By using generated questions as an extra signal, significant QA improvement can be achieved
  - [Duan et al., 2017]

- (**QG helps QA**) QG can help models achieve better QA performance using a generative machine comprehension model
  - [Wang et al., 2017]

- (**QG helps QA**) QG can improve the performance of QA over Knowledge Base (KBQA)
  - [Hu et al., 2019]

- (**QA helps QG**) Proposal of a QA-based evaluation method which measures the QG model's ability to mimic human annotators in generating QA training data
  - [Zhang and Bansal, 2019]

System overview - main steps



Left side (**red**): Question Generation
Right side (**blue**): Question Answering
Dashed line highlights the input and output elements from both modules

Question generation: syntax-based approach

- **Mainly generates factual** questions from text;
- Converts the declarative sentence target into an interrogative sentence by **manipulation of syntactic structure of the sentence**;
- Identifies the **syntactic elements** that are further used to perform the necessary transformations.

Sentence: The French Revolution *was a period of intense political and social upheaval in France.*
Question: Which event was a period of intense political and social upheaval in France?

Question generation: syntax-based approach

1. Perform **Part of Speech Tagging**;
2. Perform **Named Entity Recognition**;
3. **Build a representation** from the combination between POS and NER;
4. Use **regex** in order to match certain patterns;
5. **Transform** declarative sentence into an interrogative sentence;
6. **Formulate** the question.

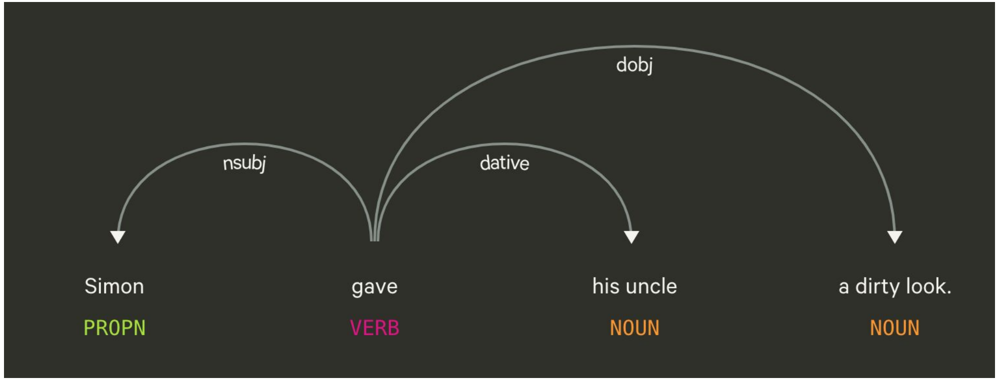
Sentence pattern before POS and NER	**The French Revolution** was a period of intense political and social upheaval in France.
Sentence pattern after POS	<DET><PROPN><PROPN><AUX><DET><NOUN><ADP><ADJ><ADJ><CCONJ><ADJ><NOUN><ADP><PROPN><PUNCT>
Sentence pattern after NER	<**EVENT**><AUX><DET><NOUN><ADP><ADJ><ADJ><CCONJ><ADJ><NOUN><ADP><PROPN><PUNCT>
Expression used as rule in QG	<**EVENT**><(?:AUX|VERB)>.*?<PUNCT>
Generated Question	**Which event** was a period of intense political and social upheaval in France?

Syntax-based Approach - More Examples

Entity/Entities	Sentence and Question
PER = Paul	S: **Paul** was the son of Henry of Burgundy and Teresa, the illegitimate daughter of King Alfonso VI of León and Castile. Q: **Who** was the son of Henry of Burgundy and Teresa?
PER = Anne PER = Henry	S: **Henry and Anne** reigned jointly as count and countess of Portugal. Q: **What people** reigned jointly as count and countess of Portugal?
GPE = Portugal	S: **Portugal** was conquered by Afonso I. Q: **Which country** was conquered by Afonso I?
ORG = The Congress of Manastir	S: **The Congress of Manastir** had chosen the Latin script as the one to be used to write the language. Q: **Which organization** had chosen the Latin script as the one to be used to write the language?
MONEY = 50 thousand dollars	S: One bedroom apartment costs **50 thousand dollars**. Q: **How much** costs the one bedroom apartment?
PER = Henry	S: A car is cleaned by **Henry**. Q: **Who** did clean a car?
DATE = 1109	S: Paul was born in **1109**. Q: **When** was Paul born?

Dependency-based Approach - What is it?

- **Connects words** in a sentence in a **graphical structure** based on their grammatical and **functional relations**;
- We can ask about **What?**, **To whom?** using direct and indirect complement.



Sentence: *Simon gave <u>his uncle</u> <u>a dirty look</u>*.

Question 1: **What** did simon give?
Question 2: **To whom** did Simon give a dirty look?

Question Analysis

- Given a **certain question**:

> *Which event was a period of intense political and social upheaval in France?*

- **Extract** all the **entities**: *France*
- **Extract** all the **subjects**: *event*
- and **objects**: *political*, *social*
- Use the **two previous steps** as **tokens**: *France*, *event*, *political*, *social*

Passage Retrieval

Question: *Which event was a period of intense political and social upheaval in France?*

- Extract **sentences** containing **at least one token**
- **Rank** the sentences using a similarity score
- Keep the sentences that have a cosine-similarity based score **higher than 75%** that will be, at most, **3 sentences**

- **Sentences**:
  1. The ***French*** Revolution was a period of intense ***political*** and ***social*** upheaval in ***France***.
  2. ***France*** tends to be a revolutionary ***political*** country
  3. *Trés bué* is an ***event*** happening in ***France***.

Answer Extraction

Question: *Which event was a period of intense political and social upheaval in France?*

Extracted Triples:

- **Extract all triples** from the sentences
- Discard all the triples in which their **relation is not in** the question
- If it only contains the subject or object **the remaining element will be considered the answer**. Otherwise, **no answer** is selected

- {'subject': '**France**' , 'relation': '**tends to be**', 'object': '**revolutionary**'}
- {'subject': '**French Revolution**', 'relation': '**was period of**', 'object': '**intense political upheaval in France**'}
- {'subject': '**Revolution**', 'relation': '**was period of**', 'object': '**political upheaval in France**'}
- {'subject': '**French Revolution**', 'relation': '**was**', 'object': '**period**'}

1. Evaluation (QG)

- *How to evaluate the **quality** of the computer-generated questions?*

2. Evaluation (QA)

- *How to evaluate the **accuracy** of the given answers?*

Dataset

- ***Wiki2sents***:
  - collection of **7.8 million sentences** from August 2018 English Wikipedia dump;

- Documents created:
  - **[Dataset 1] Wiki Documents**: **1000 sentences** were randomly extracted from *wiki2sents;*
  - **[Dataset 2] Controlled Document**: human selected sentences to handle QG specific cases.

QG - A Survey with Teachers

- Creation of a survey with **10 generated questions**;
- **4 metrics per generated question**;
  - **Objectivity** of the Question - “*Do you consider the question objective?*”
    - (1- Nothing objective, 5 - Very objective)
  - **Grammatically** - “*Do you consider the question to be grammatically correct?*”
    - (1 - Very Incorrect, 5 - Totally Correct)
  - **Question Extension** - “*Do you consider the extension/length of the question adequate?*”
    - (1 - Not too long, 5 - Too long)
  - **Answer** - “*How many answers do you think this question might have?*”
    - (No answer, One, Two or more)
- **5 English Teachers** answered the created survey.

QG - Results: Objectivity and Grammaticality

- Objectivity = **3,64** (1...5)
  - Overall the Teachers **found the questions objective**;
  - Ambiguities aroused in questions could occur when there are **multiple entities**;
  - Some questions can be **too generic**.

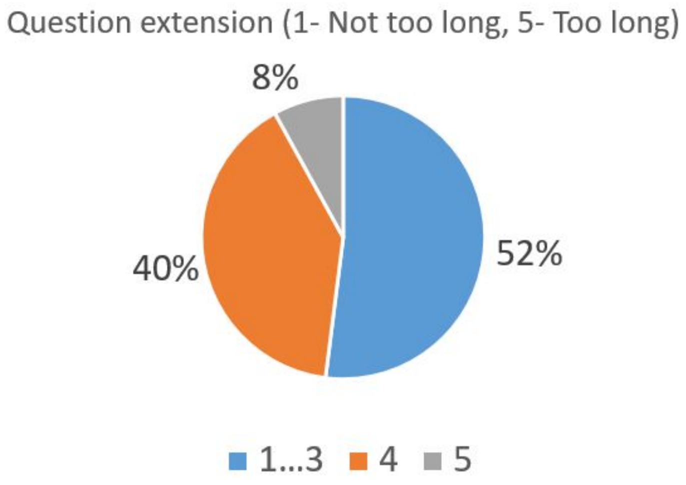
- Grammaticality = **3,42** (1...5)
  - The **questionable term** used at the beginning of the sentence may not be the most appropriate;
  - Main faults from **the conjugation of verbs**.

QG - Results: Question Extension

- The size of the question is **adequate most of the time**;

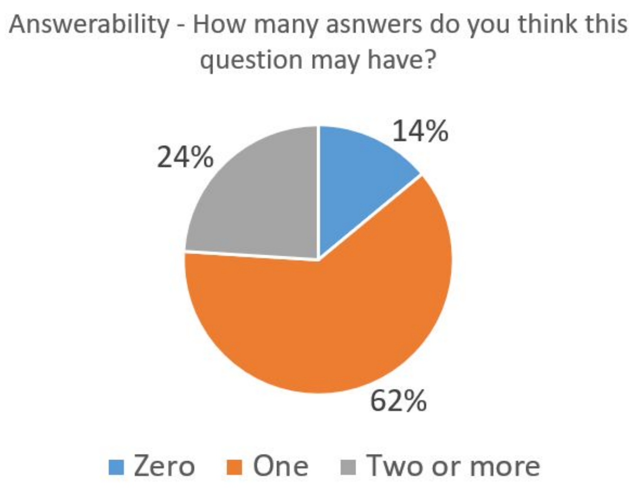
- Needs more treatment to **remove unnecessary parts**;

- The appropriateness of the question length may **depend on the type of content** that is questioned.



QG - Results: Answerability

- **Little consensus** regarding the number of answers given a question;
- There are **several interpretations**;
- This can be caused by the presence of **multiple entities** in the sentence or **external knowledge** (in addition to what is written in the sentence).



QA - Results

- **Short Answer**: Mean confidence between correct answer and predicted answer
- **Correct Triple**: 100% if the triple contains the answer, otherwise 0
- **Correct Sentence**: 100% if the answer is in the found sentence, otherwise 0

Dataset	Question Type	Number of generated questions	Avg. Short Answers	Avg. Correct Triples	Avg. Correct Sentences
Dataset 1: Wiki Documents	Entities + Dependency	311	78,5%	87,4%	96,4%
Dataset 2: Controlled Document	Entities + Dependencies	43	89,3%	95,2%	100%
Dataset 1 + 2	Entities without ORG	301	80,9%	88,5%	98,8%
Dataset 1 + 2	Entities including ORG	334	**81,2%**	89,8%	98,1%
Dataset 1 + 2	Dependencies	20	**32,7%**	80,0%	100%

Conclusions

- QA can help QG by **detecting ambiguous** questions
- QG can help QA by **generating data sets** with question and answer pairs in a **completely automatic way** -> **less human intervention** will be necessary to create this type of content

- Future work:
  - Create questions by chaining different sentences;
  - In a learning perspective, create datasets whose content is from **different subjects;**
  - **Evaluate** state-of-the-art **QA models** by using, as input, different questions from each **QG technique.**

Thank you!

P. Azevedo, B. Leite, H. L. Cardoso, D. C. Silva, L. P. Reis
Artificial Intelligence and Computer Science Lab
University of Porto, Portugal

AIAI 2020 - 16th International Conference on
Artificial Intelligence Applications and Innovations
5 – 7 June, 2020

References

- *Ezz, Ahmed & Dahab, Mohamed. Automatic Generation of Question Bank Based on Pre-defined Templates. 10.13140/2.1.3633.3766, 2014.*
- *Le, Nguyen-Thinh & Pinkwart, Niels. (2015). Evaluation of a question generation approach using semantic web for supporting argumentation. Research and Practice in Technology Enhanced Learning. 10. 10.1007/s41039-015-0003-3.*
- *Majumder, Mukta & Saha, Sujan Kumar. (2015). A System for Generating Multiple Choice Questions: With a Novel Approach for Sentence Selection. 64-72. 10.18653/v1/W15-4410.*
- *Guy Danon and Mark Last. A syntactic approach to domain-specific automatic question generation. CoRR, abs/1712.09827, 2017.*
- *Flor, Michael & Riordan, Brian. A Semantic Role-based Approach to Open-Domain Automatic Question Generation. 254-263. 10.18653/v1/W18-0530, 2018.*
- *Jun Araki, Dheeraj Rajagopal, Sreecharan Sankaranarayanan, Susan Holm, Yukari Yamakawa, and Teruko Mitamura. Generating questions and multiple-choice answers using semantic analysis of texts. In COLING. pages 1125–1136, 2016.*
- *Mazidi, Karen & Tarau, Paul. (2016). Infusing NLU into Automatic Question Generation. 51-60. 10.18653/v1/W16-6609.*
- *CHEN, G.; YANG, J.; HAUFF, C.; HOUBEN, G.. LearningQ: A Large-Scale Dataset for Educational Question Generation. International AAAI Conference on Web and Social Media, North America, jun. 2018*
- *Lu, Xinyuan. (2019). Learning to Generate Questions with Adaptive Copying Neural Networks. 1838-1840. 10.1145/3299869.3300100.*
- *Kun Xu, Sheng Zhang, Yansong Feng, and Dongyan Zhao. Answering natural language questions via phrasal semantic parsing. In Natural Language Processing and Chinese Computing, pages 333–344. Springer, 2014.*
- *Abraham Ittycheriah and Salim Roukos. Ibm's statistical question answering system-trec-11. Technical report, IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS NY, 2006.*
- *Abdessamad Echihabi, Ulf Hermjakob, Eduard Hovy, Daniel Marcu, Eric Melz, and Deepak Ravichandran. How to select an answer string? In Advances in open domain question answering, pages 383–406. Springer, 2008.*
- *Partha Pakray, Pinaki Bhaskar, Somnath Banerjee, Bidhan Chandra Pal, Sivaji Bandyopadhyay, and Alexander F Gelbukh. A hybrid question answering system based on information retrieval and answer validation. In CLEF (Notebook Papers/Labs/Workshop), 2011*
- *Rob High. The era of cognitive systems: An inside look at ibm watson and how it works. IBM Corporation, Redbooks, 2012.*
- *Saeedeh Shekarpour, Edgard Marx, Axel-Cyrille Ngonga Ngomo, and SA Sina. Semantic interpretation of user queries for question answering on interlinked data Elsevier-Web Semantics, 2015.*
- *Hu, Sen & Zou, Lei & Zhu, Zhanxing. (2019). How Question Generation Can Help Question Answering over Knowledge Base. 10.1007/978-3-030-32233-5_7.*
- *Duan, Nan & Tang, Duyu & Chen, Peng & Zhou, Ming. (2017). Question Generation for Question Answering. 866-874. 10.18653/v1/D17-1090.*
- *Zhang, Shiyue & Bansal, Mohit. (2019). Addressing Semantic Drift in Question Generation for Semi-Supervised Question Answering.*
- *Tong Wang, Xingdi Yuan, Adam Trischler (2017). A Joint Model for Question Answering and Question Generation.*