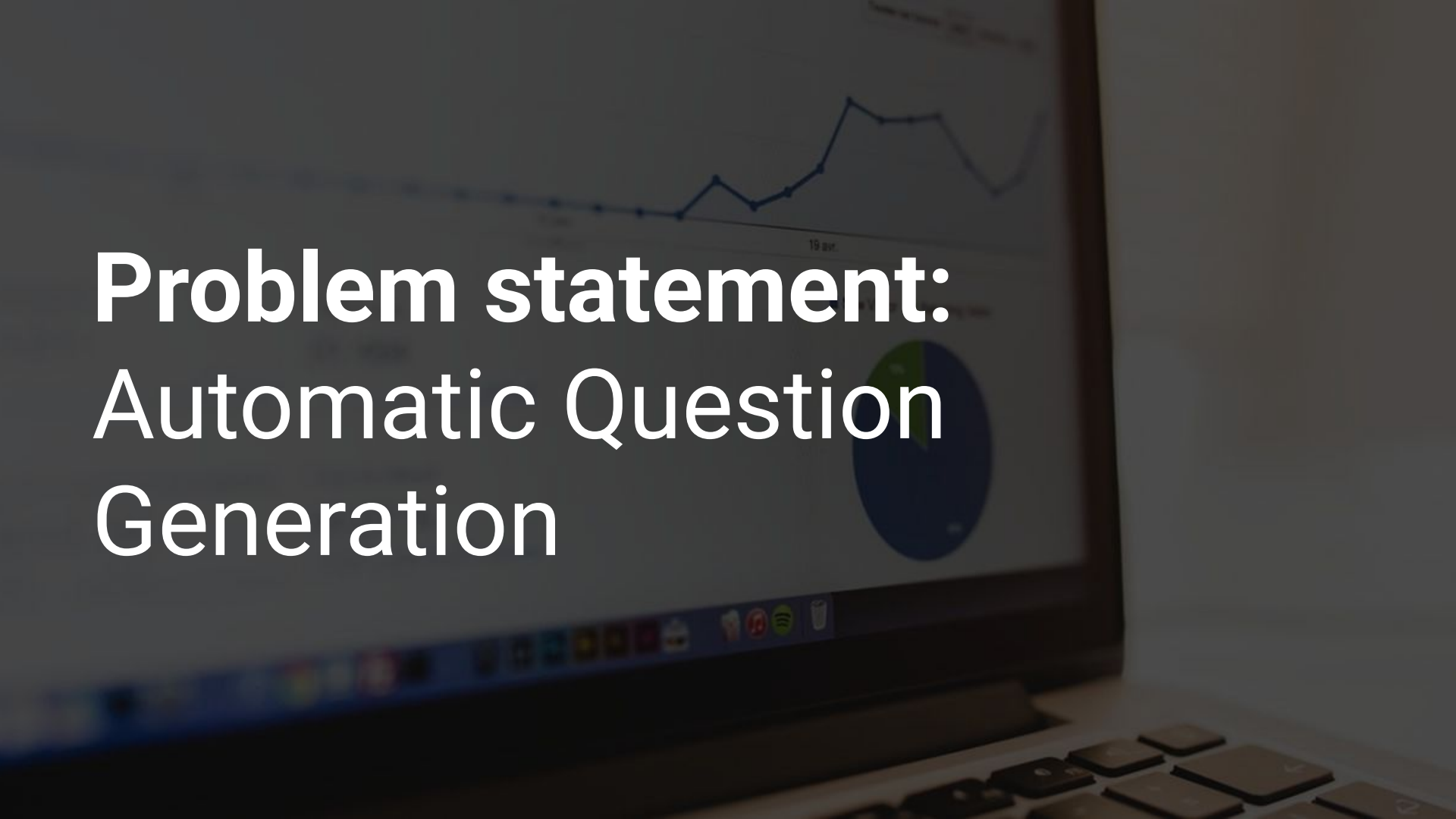


Embibe

Gaand Maar Lenge



The background image shows a laptop screen with a dark overlay. On the screen, there is a line graph with a blue line and a pie chart with blue and green segments. The text 'Problem statement: Automatic Question Generation' is written in white, bold, sans-serif font across the center of the screen. The laptop keyboard is visible at the bottom right.

Problem statement: Automatic Question Generation

The problem

You are given a set of books, you need to generate the questions out of that for different grades and levels. Possible types of question can be any of the following:

- Fill in the blanks
 - Question that starts with W/H
 - MCQ type questions
 - Generic questions (genericity depends on the subjects)
-



The solution

PREPROCESSING

The data is firstly preprocessed to remove incomplete sentences, new line gaps and figures.

A close-up photograph of a person's hand, wearing a dark sleeve, pointing with their index finger at a document on a desk. A pen lies on the desk near the hand. The background is blurred, showing some bokeh lights.

The solution

SENTENCE SELECTION

Weighs relevance of each sentence in the context by generating a TF/IDF score, and using heuristics like presence of nouns / pronouns / adverbs and length of sentences.

A close-up photograph of a person's hand holding a pen, poised to write on a document. The background is blurred, showing some bokeh lights. The text 'The solution' is overlaid in white on the left side of the image.

The solution

FILL IN THE BLANKS - 1

Rule based approach.
Select sentences which
have a pattern of Parts of
Speech occurring and
replace the word which
gives the most sensible
question.



The solution

FILL IN THE BLANKS - 2

Named Entity Recognition.
The tokens within the sentences are grouped into more abstract classes.



The solution

FILL IN THE BLANKS - 2

After being classified into POS , the Named Words are then chosen on random to replace with the blank. This generates a better result than only POS.

A close-up photograph of a hand holding a pen, poised to write on a document. The background is blurred, showing what appears to be a person's face in profile. The text 'The solution' is overlaid in white on the left side of the image.

The solution

FILL IN THE BLANKS - 2

After being classified into POS , the Named Words are then chosen on random to replace with the blank. This generates a better result than only POS.

A close-up photograph of a person's hand holding a pen, writing on a piece of paper. The background is blurred, showing some bokeh lights. The text 'The solution' is overlaid in white on the left side of the image.

The solution

WH Questions Generation-1

The n-gram Language Model is generated from the training set. A first n-gram is picked by a random choice from a set. Next a random number between 0 to 1 is generated and the corresponding n-gram is picked up till a question mark is encountered.

W/H Questions: Approach 1

- Overgenerate and then rank the questions through a regression model to find the most optimal questions out of those.
- Part 1: Overgeneration
 - First, take a complicated sentence and change it into simpler, derived sentences using lexical items, syntactic structures, compression, etc.
 - Then, take the derived sentences and use a rule based approach to convert them into W/H questions.

W/H Questions: Approach 1

- Part 2: Ranking questions using a regression model
 - We can create datasets through teachers to rank the questions, and run it on our generated questions to predict a rank.
 - Then we can select the top questions based on those predictions.

MCQ Type: Approach 1

- Part 2: Ranking questions using a regression model
 - We can create datasets through teachers to rank the questions, and run it on our generated questions to predict a rank.
 - Then we can select the top questions based on those predictions.

W/H Questions: Approach 3

- Specific patterns of a POS sequence are pre-loaded.
- Sentences which have that combination are selected and the question is generated substituting the given Part of Speeches from the sentences words.

The team

Answer the question, "Why are we the ones to solve the problem we identified?"



Wendy Writer

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor



Ronny Reader

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor



Abby Author

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor

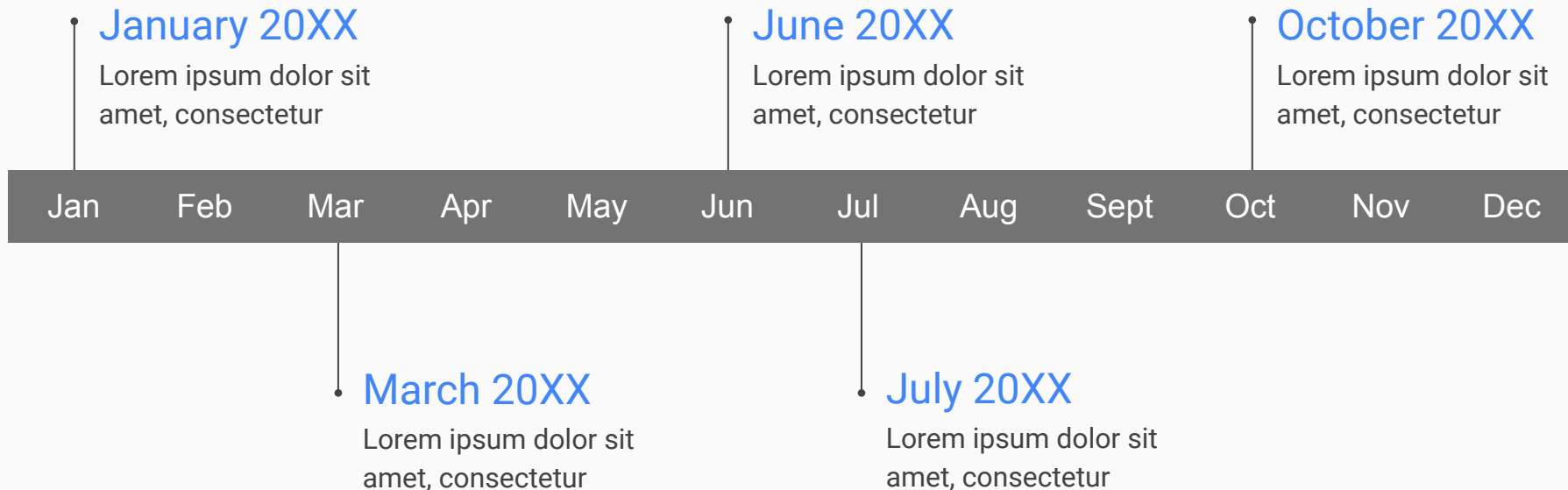


Berry Books

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor

Milestones

Show where you are in the process and what's left to tackle



Appendix

Show the audience you anticipated their questions.

Leave room for Q&A, but use the Appendix as a way to show that you both thought about those questions and have solid answers with supporting information. Let the audience test their understanding of the problem and the solution you've outlined - questions give them a chance to talk themselves into your approach, and give you a chance to show mastery of the subject.

How it works

Step 1

Lorem ipsum dolor sit
amet, consectetur



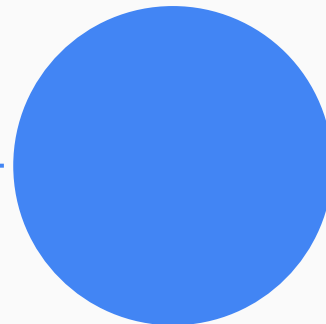
Step 2

Lorem ipsum dolor sit
amet, consectetur



Step 3

Lorem ipsum dolor sit
amet, consectetur

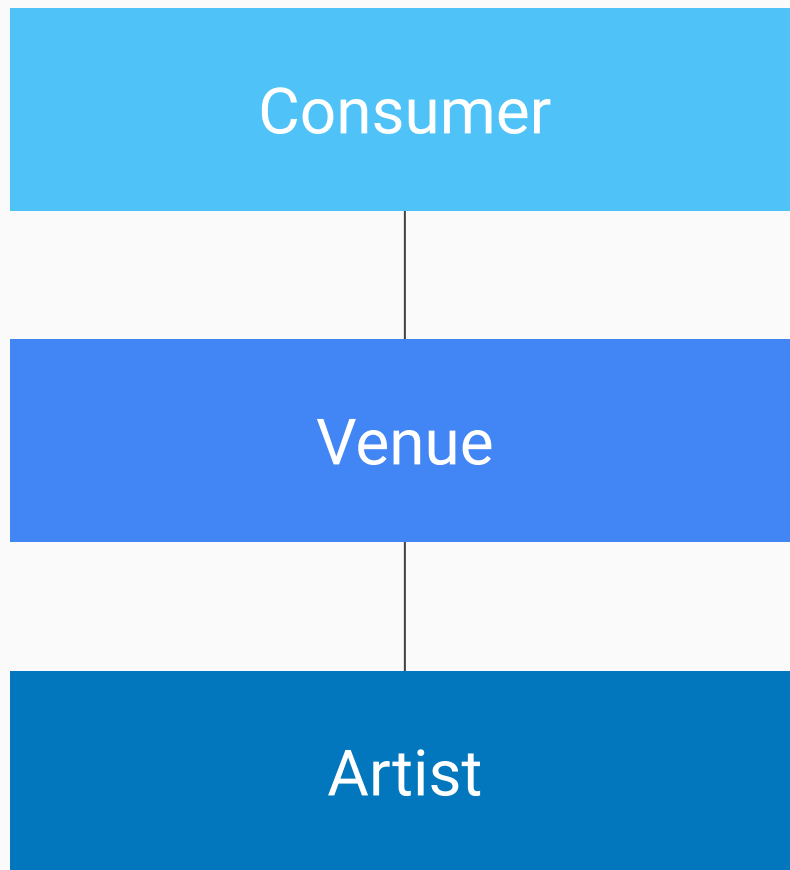


An aerial photograph of a city skyline at dusk or dawn. The sky is a mix of dark blue, purple, and orange. The city is densely packed with skyscrapers, many of which are illuminated with lights. The text "The technology: GPS + RFID" is overlaid in a large, white, sans-serif font on the left side of the image.

The technology: GPS + RFID

Revenue model

Lorem ipsum dolor sit amet,
consectetur adipiscing elit,
sed do eiusmod tempor
incididunt



Why now?

Lorem ipsum dolor sit
amet, consectetur
adipiscing elit, sed do
eiusmod tempor
incidunt

