Team Project Proposal

Generate Captions of Images In Your Documents

Service Oriented Computing
2019 Vu Anh Van
20194334 Soyeon Kim

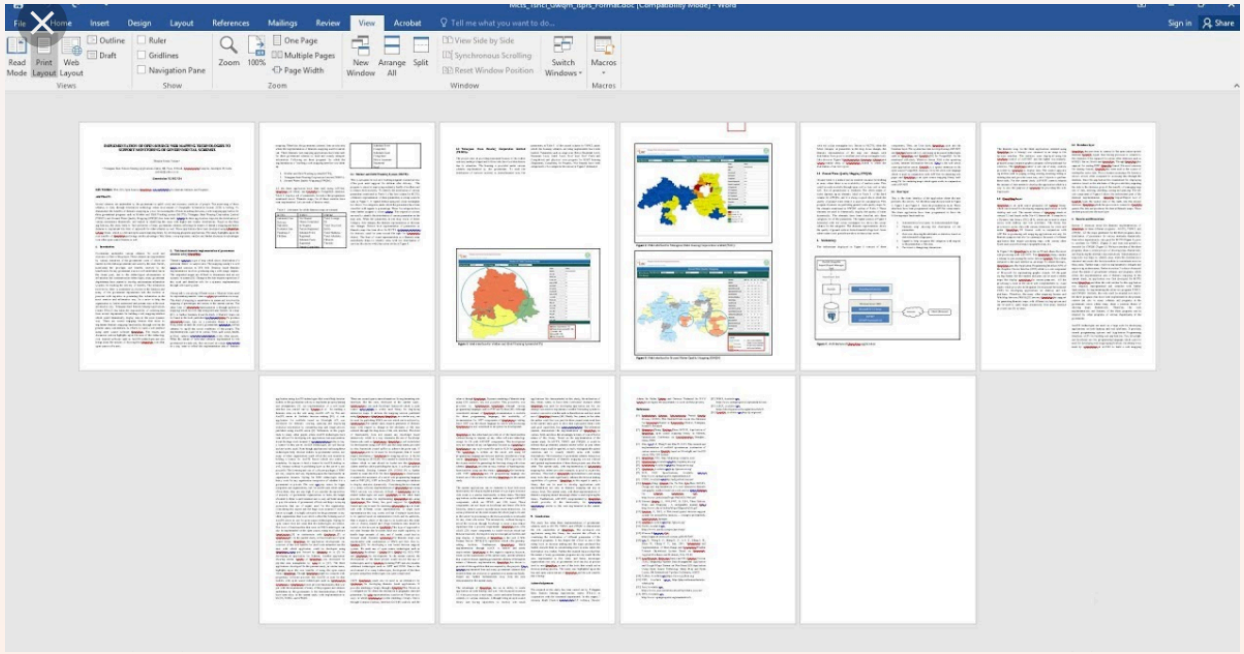
KAIST

Contents

KAIST

➢ Background

- Usually we put multiple images in our documents
- Tedious works to write each caption of each images in documents

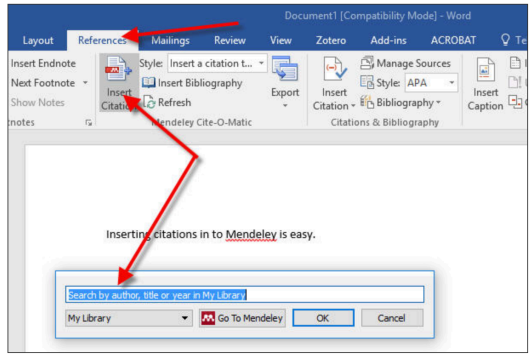


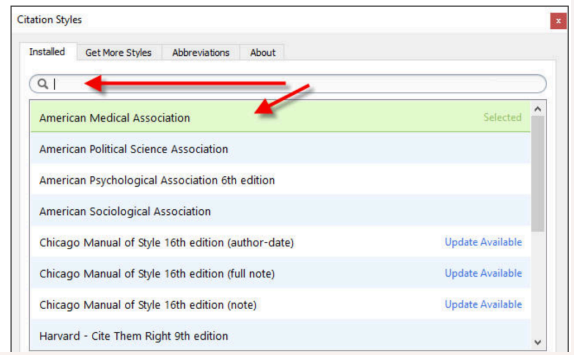
What about a service that generates captions automatically?

➢ Benchmark service

- There are kinds of services that generate a certain format of citation for papers

Ex. Mendeley





➢ Generate proper captions for images in documents automatically

- Image captioning: automatically generating a natural language description of an image.
- Captions are usually related to the not only images but also the sentences in the documents → computer vision & NLP

➢ Make a prototype as a web service

- The service integrated this function would be nice with Microsoft office and so on
- But, first we'll make a prototype as a web service

➢ Generate proper captions for images in documents

- Image captioning
  - automatically generating a natural language description of an image
  - Very related to computer vision and natural langue processing

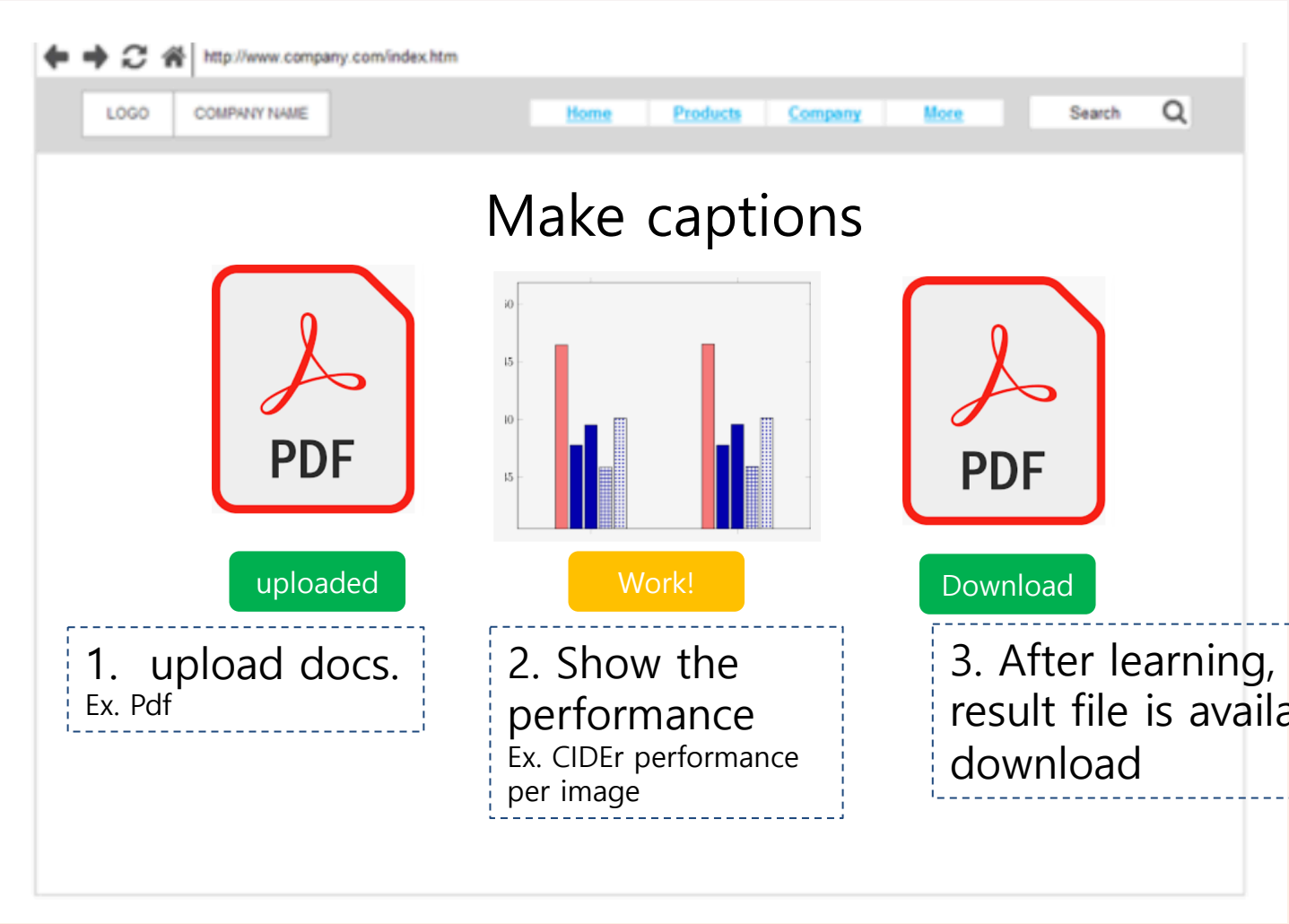
"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

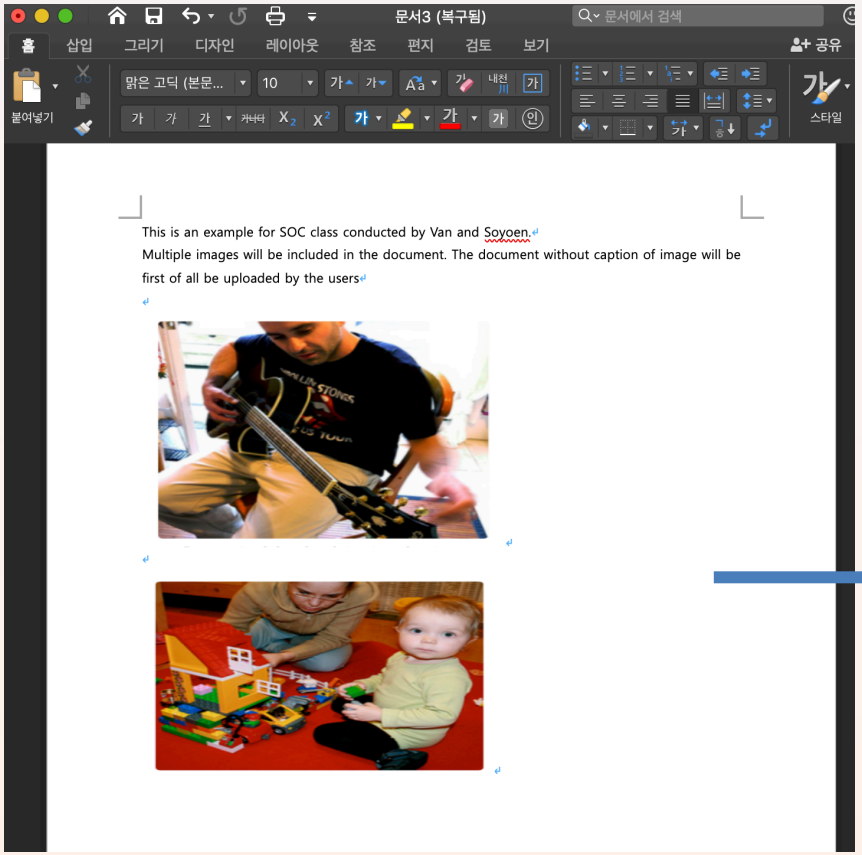
KAIST

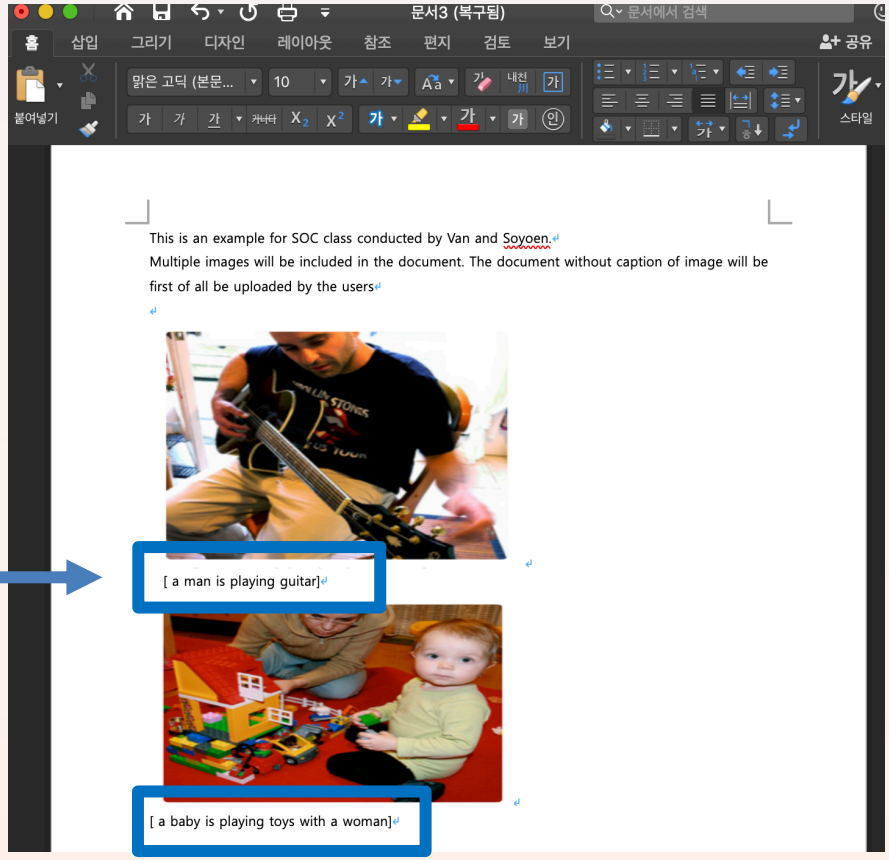
➢ Simple draft of a service look



Proposed Idea

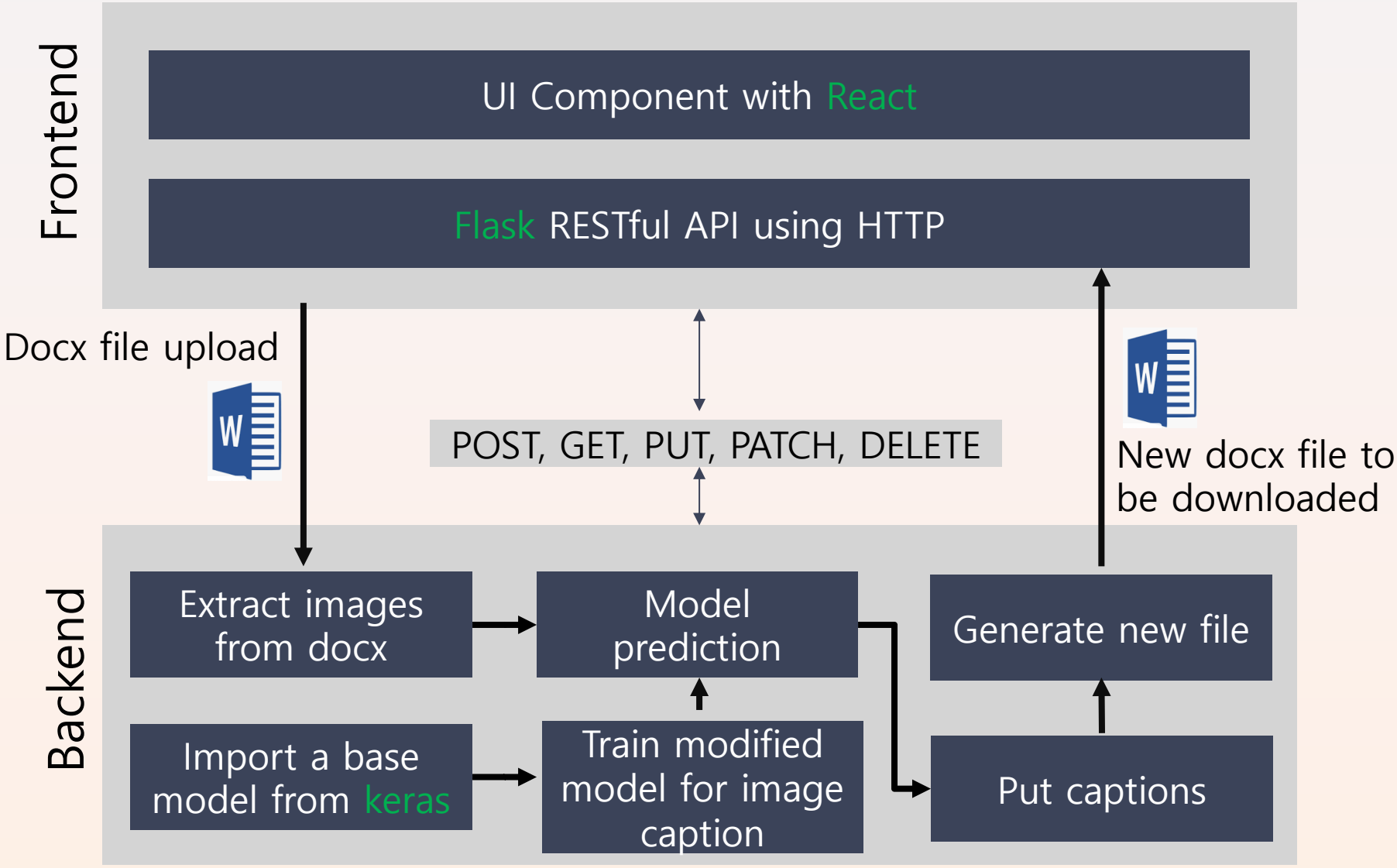
➢ If upload the file, the expected results are …





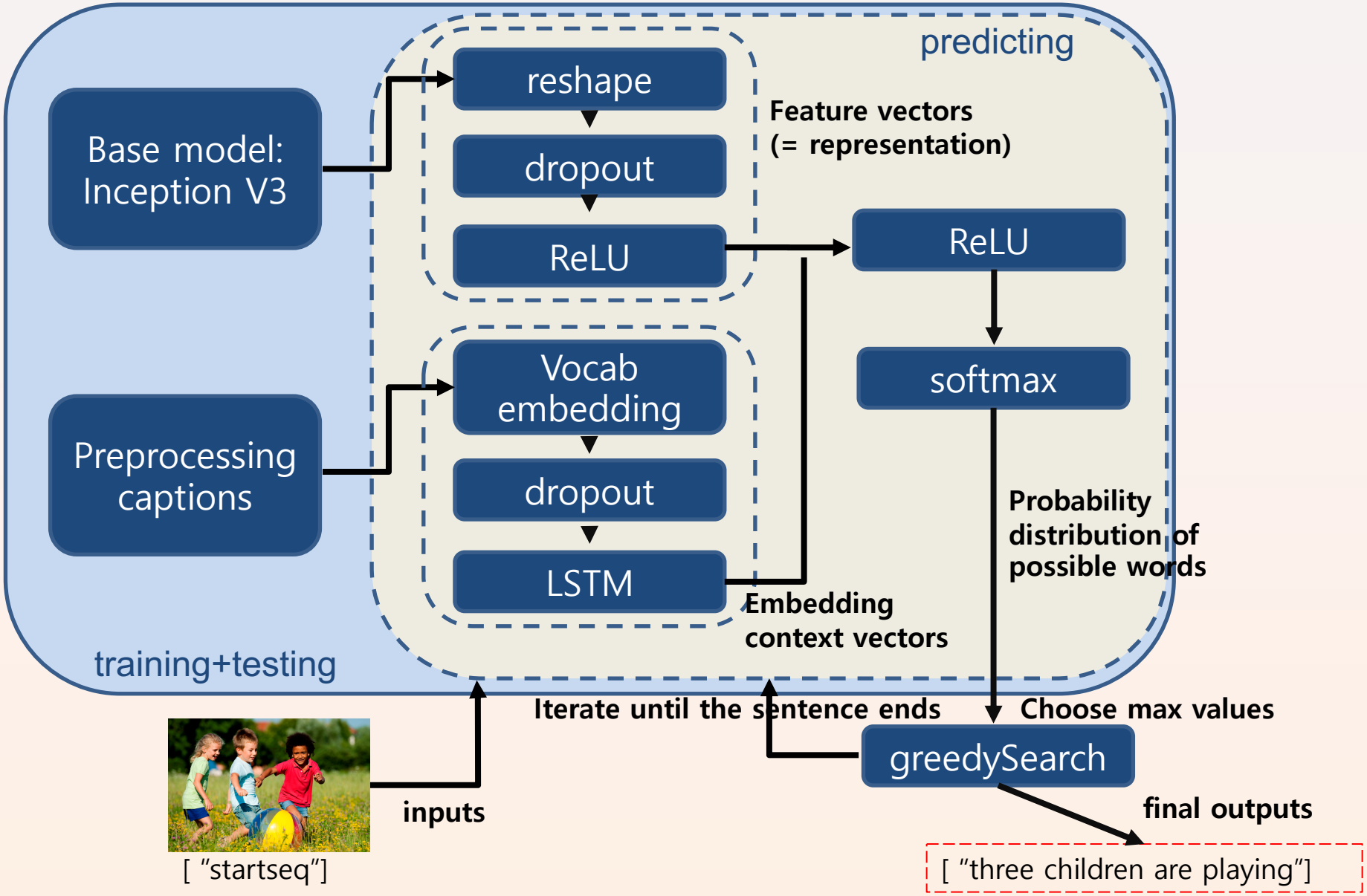
KAIST

➢ Overall Service Architecture



KAIST

➢ Overall Deep learning model Architecture



KAIST

➢ Training model summary

- Training/test dataset: Flickr8k
  - Images with 5 different possible captions
  - Training: 6000 , test: 1000
- 2 main deep learning models
  - **Convolutional Neural Network(CNN)**: to catch overall feature of images
  - **Long Short Term Memory(LSTM):** to produce full sentences based on the prediction using each part of images and which words it produced so far
- Loss function: crossentropy
- Optimizer: Adam
- Learning rate=0.001
- Epochs=10
- CPU

➢ Web deployment & frameworks

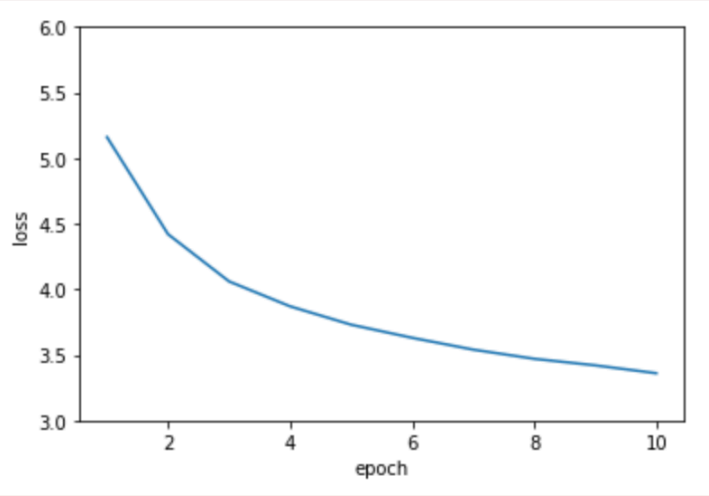
- Flask, React.js
- Keras

➢ Expected Results & Performance analysis(until next week)

- BLEU score for texts which seem plausible
- Loss

Results

➢ Performance: loss



End loss: 3.3674

➢ BLEU score: an algorithm for evaluating the quality of text which has been machine translated

```
In [16]: prediction=["white dog is running through the grass",
                     "boy in red shirt is playing soccer ball",
                     "woman walking down sidewalk street",
                     "two dogs play in the grass",
                     "two people are sitting on the street"]
         target=["two white dogs are running in the grass",
                 "a boy wearing a red shirt is playing soccer",
                 "A woman and a man are walking down the street",
                 "two dogs play in the grass",
                 "four people are having a party on the street"
                ]
         bleu=get_bleu(prediction, target)

In [17]: print(bleu)

         67.48494264632554
```

Average BLEU score: 67.49

Results

- Demo!

Thank you