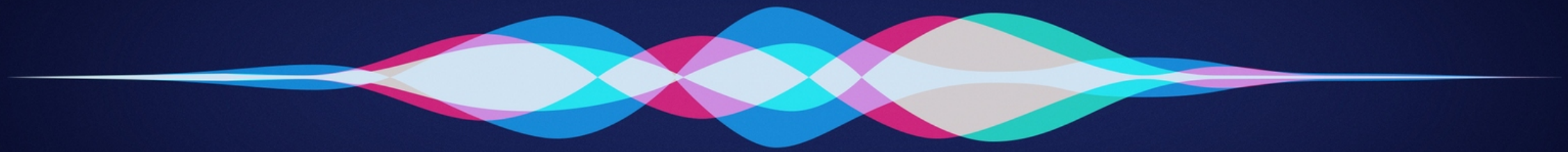
Question Answering System: JARDIS



Week 4: Memory Network

Zakaria

김준식 류영표 손규빈
유재준 이현호 정유진

Goal – Study – Plan

Question answering system
for **specific** domain

Stack Overflow?

조선왕조실록?

Movie script?

Travel information?

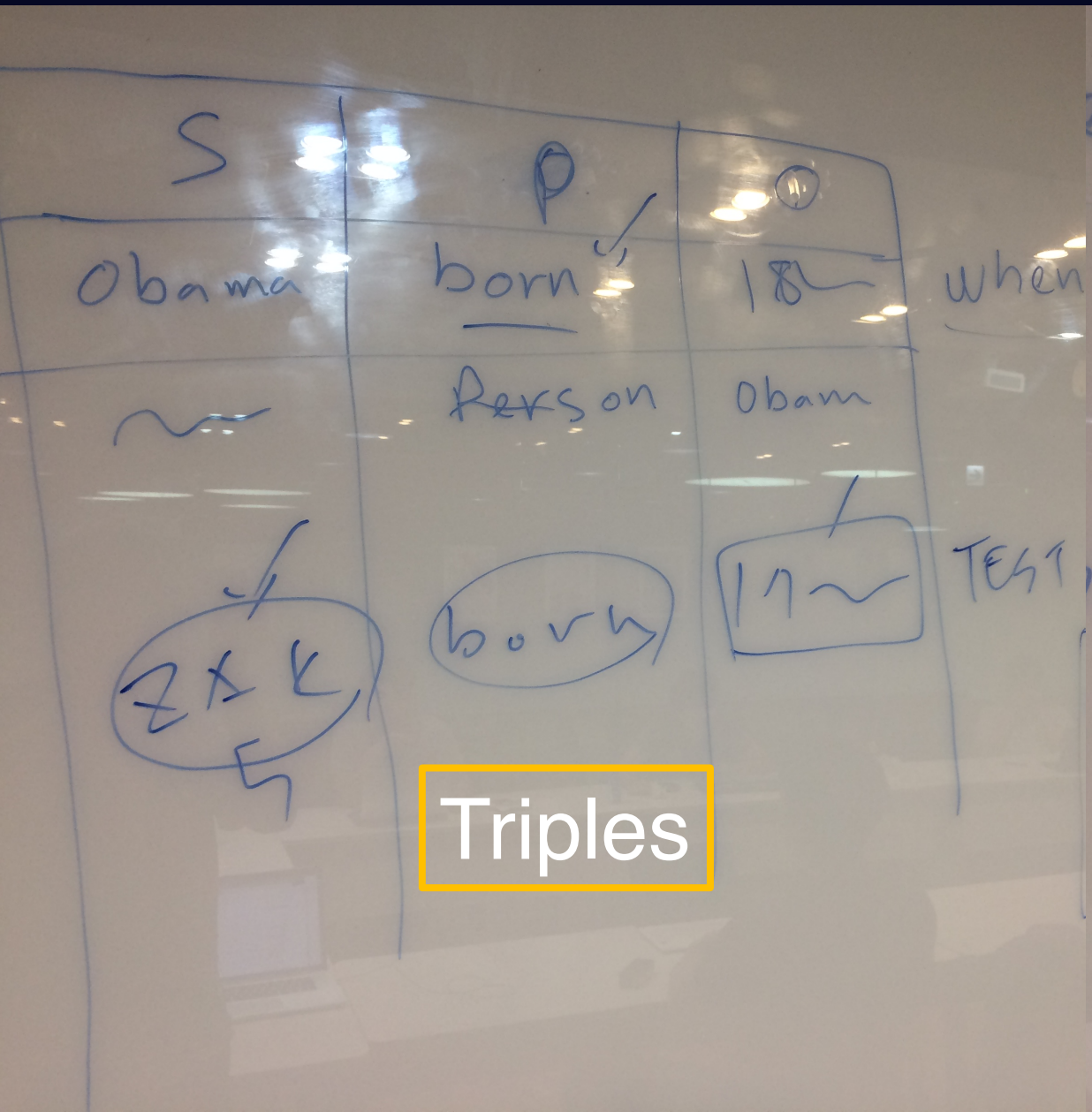
한글?

Goal – Study – Plan

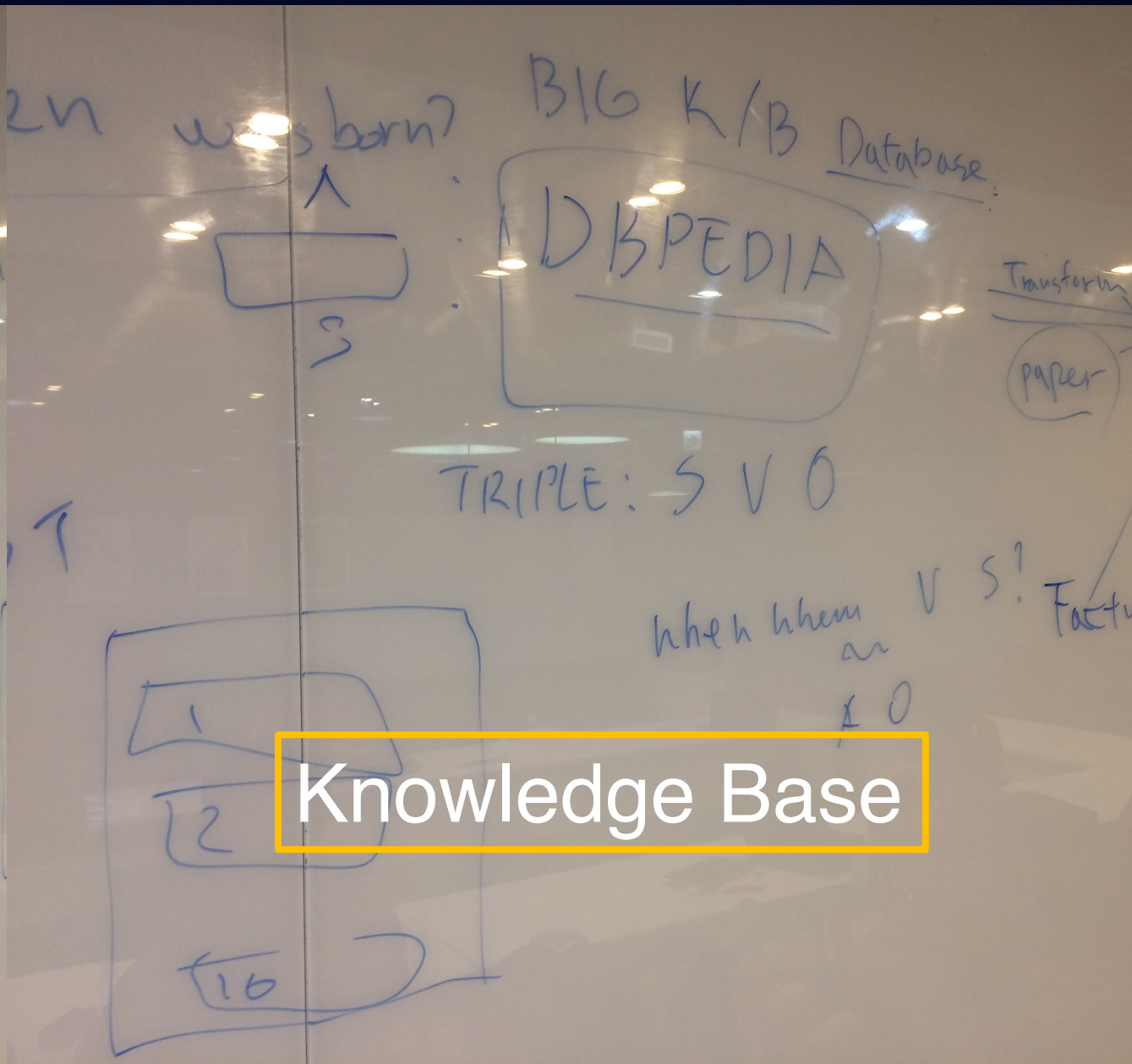
W1: Word embedding	Word2Vec(Skip-gram, CBOW), Gensim, GloVe
W2: Sequence embedding	RNN, GRU, LSTM, Seq2Seq
W3: Question Answering	Question Answering Using Deep Learning, Deep Averaging Network(+video), Question Answering with Subgraph Embeddings
W4: Memory Network	Memory Networks(+video), End-To-End Memory Networks, Dynamic Memory Networks for Visual and Textual Question Answering
Goal setting, Dataset	Not yet
QA Implementation	Not yet

Goal – Study – Plan

Goal – Study – Plan

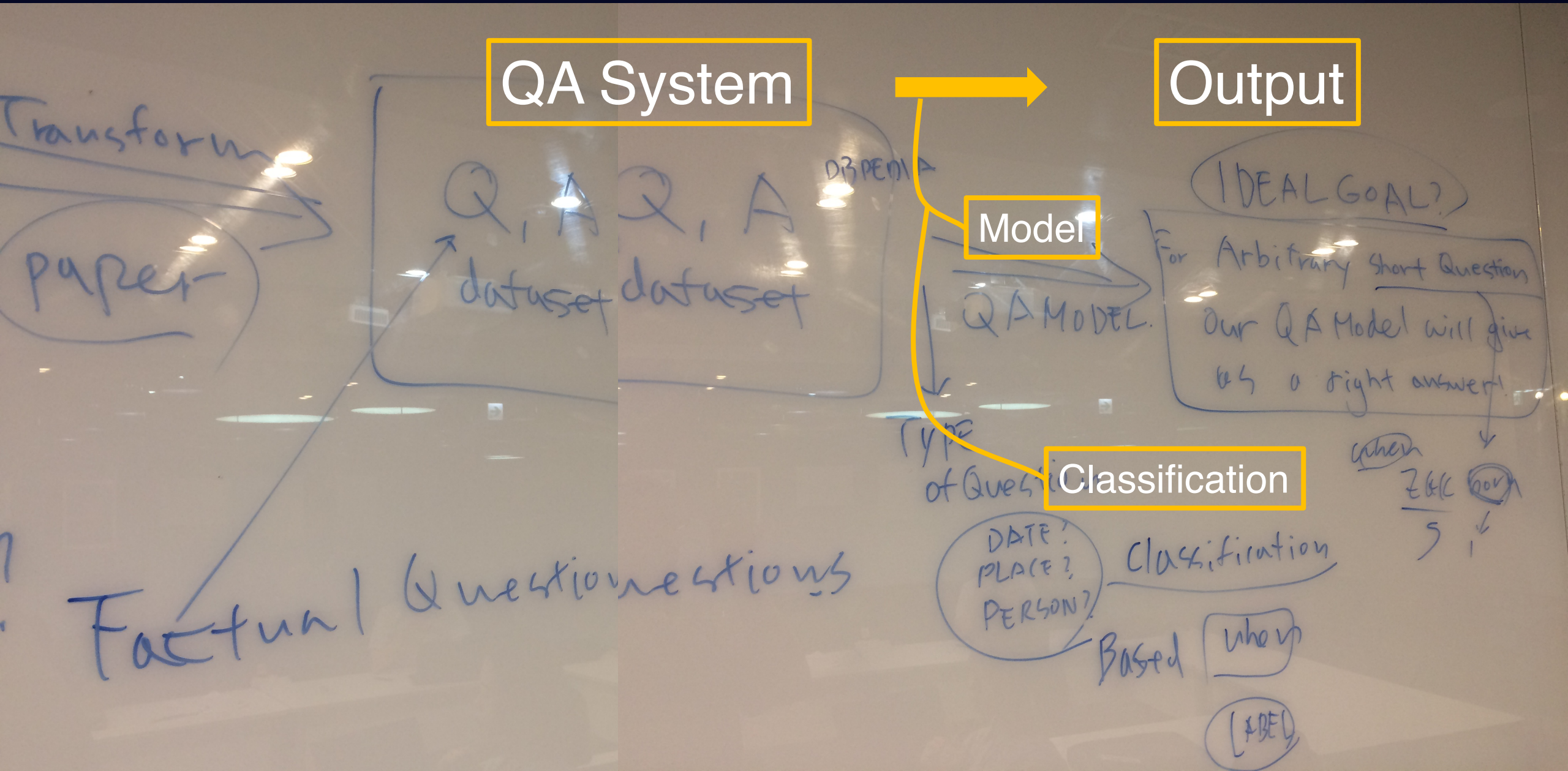


Triples



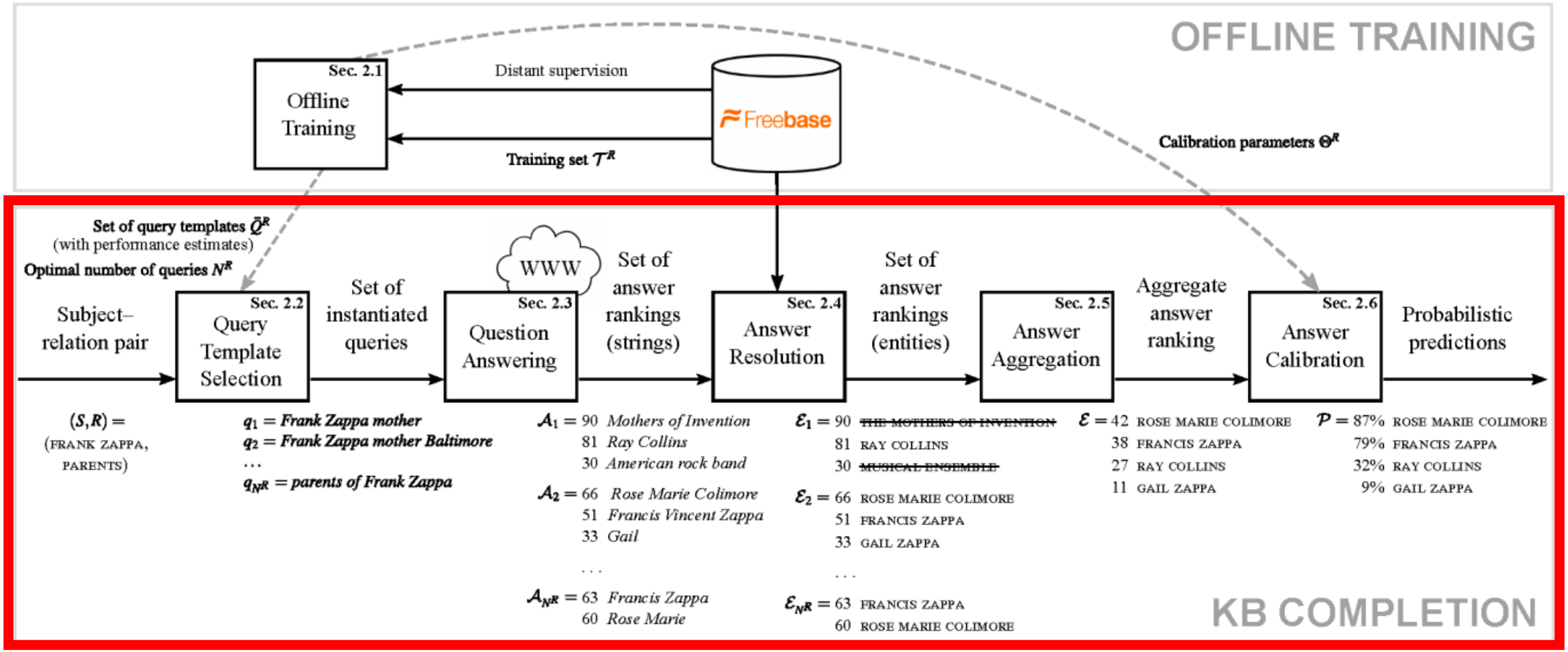
Knowledge Base

Goal – Study – Plan



Goal – Study – Plan

"Knowledge Base Completion via Search-Based Question Answering" by B.West et al 2014 (WWW)



Memory Network

Published as a conference paper at ICLR 2015

youtu.be/Xumy3Yjq4zk

MEMORY NETWORKS

Jason Weston, Sumit Chopra & Antoine Bordes
Facebook AI Research
770 Broadway
New York, USA
{jase,spchopra,abordes}@fb.com

ABSTRACT

We describe a new class of learning models called *memory networks*. Memory networks reason with inference components combined with a long-term memory component; they learn how to use these jointly. The long-term memory can be read and written to, with the goal of using it for prediction. We investigate these models in the context of question answering (QA) where the long-term memory effectively acts as a (dynamic) knowledge base, and the output is a textual response. We evaluate them on a large-scale QA task, and a smaller, but more complex, toy task generated from a simulated world. In the latter, we show the reasoning power of such models by chaining multiple supporting sentences to answer questions that require understanding the intension of verbs.

End-To-End Memory Networks

Sainbayar Sukhbaatar
Dept. of Computer Science
Courant Institute, New York University
sainbar@cs.nyu.edu

Arthur Szlam **Jason Weston** **Rob Fergus**
Facebook AI Research
New York
{aszlam,jase,robfergus}@fb.com

Abstract

We introduce a neural network with a recurrent attention model over a possibly large external memory. The architecture is a form of Memory Network [23]

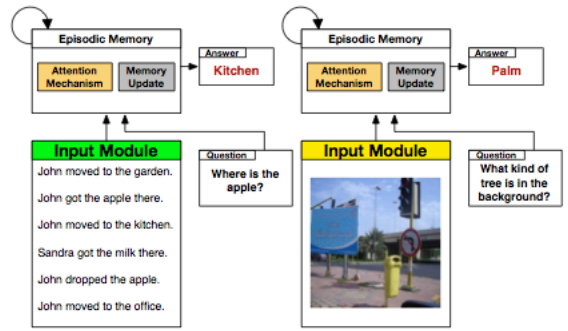
Dynamic Memory Networks for Visual and Textual Question Answering

Caiming Xiong*, Stephen Merity*, Richard Socher
MetaMind, Palo Alto, CA USA

{CMXIONG,SMERITY,RICHARD}METAMIND.IO
*indicates equal contribution.

Abstract

Neural network architectures with memory and attention mechanisms exhibit certain reasoning capabilities required for question answering. One such architecture, the dynamic memory network (DMN), obtained high accuracy on a variety of language tasks. However, it was not shown whether the architecture achieves strong results for question answering when supporting facts are not marked during training or whether it could be applied to other modalities such as images. Based on an analysis of the DMN, we propose



(a) Text Question-Answering (b) Visual Question-Answering

Memory Network

Background

- Knowledge Base data is HUGE!
- Long-term memory is required
- To remember previous dialog (short- and long-term), and respond
- Most ML has limited memory(RNN, LSTM)

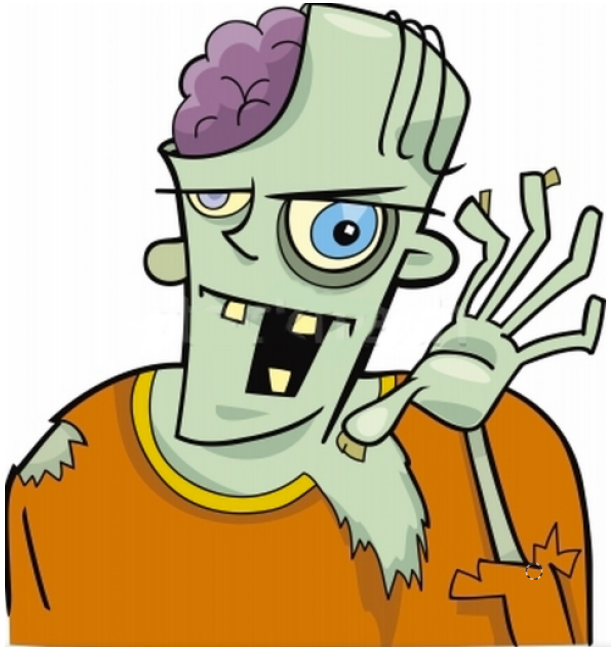
Goal – Study – Plan

Long-Term Memories	h_i	Shaolin Soccer directed_by Stephen Chow Shaolin Soccer written_by Stephen Chow Shaolin Soccer starred_actors Stephen Chow Shaolin Soccer release_year 2001 Shaolin Soccer has_genre comedy Shaolin Soccer has_tags martial arts, kung fu soccer, stephen chow Kung Fu Hustle directed_by Stephen Chow Kung Fu Hustle written_by Stephen Chow Kung Fu Hustle starred_actors Stephen Chow Kung Fu Hustle has_genre comedy action Kung Fu Hustle has_imdb_votes famous Kung Fu Hustle has_tags comedy, action, martial arts, kung fu, china, soccer, hong kong, stephen chow The God of Cookery directed_by Stephen Chow The God of Cookery written_by Stephen Chow The God of Cookery starred_actors Stephen Chow The God of Cookery has_tags hong kong Stephen Chow From Beijing with Love directed_by Stephen Chow From Beijing with Love written_by Stephen Chow From Beijing with Love starred_actors Stephen Chow, Anita Yuen . . . <and more> . . .
Short-Term Memories	c_1^u c_1^r	1) I'm looking a fun comedy to watch tonight, any ideas? 2) Have you seen Shaolin Soccer? That was zany and great.. really funny but in a whacky way.
Input	c_2^u	3) Yes! Shaolin Soccer and Kung Fu Hustle are so good I really need to find some more Stephen Chow films I feel like there is more awesomeness out there that I haven't discovered yet ...
Output	y	4) God of Cookery is pretty great, one of his mid 90's hong kong martial art comedies.

Memory Network

MemNNs have four component networks.

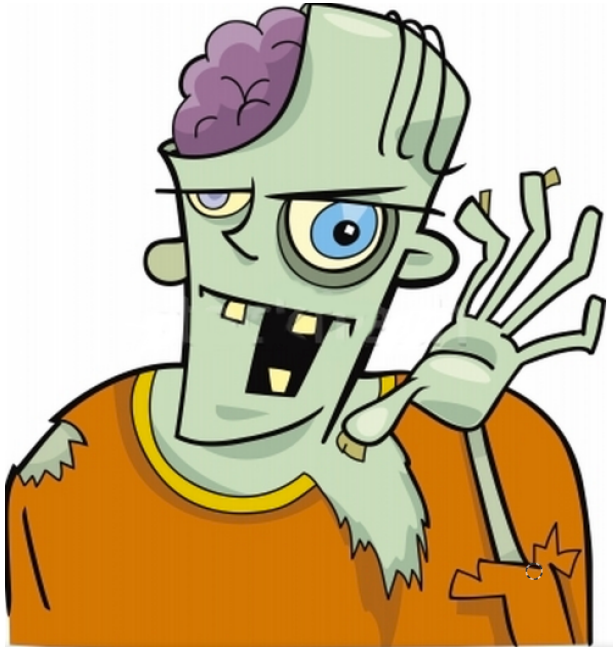
- **I** (Input feature map)
  - convert incoming data to the internal feature representation.
- **G** (Generalization)
  - update memories given new input.
- **O** (Output)
  - produce new output (in feature representation space) given the memories.
- **R** (Response)
  - convert output O into a response seen by the outside world.



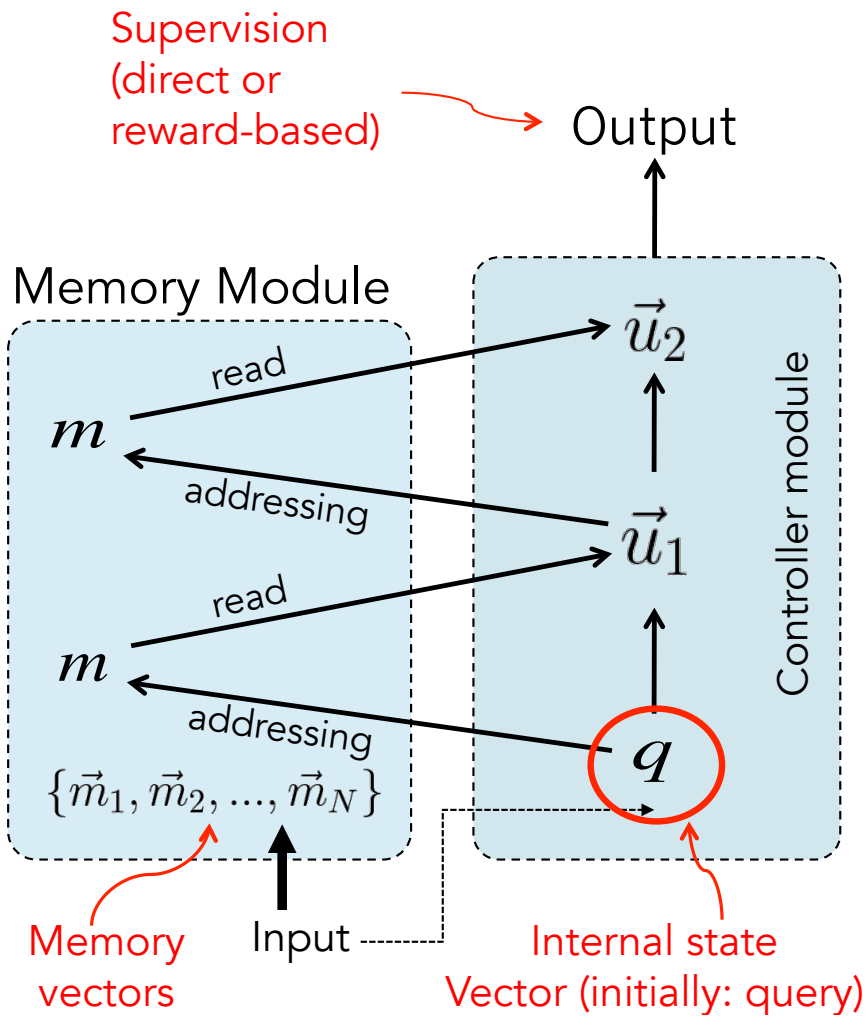
Memory Network

MemNNs have four component networks.

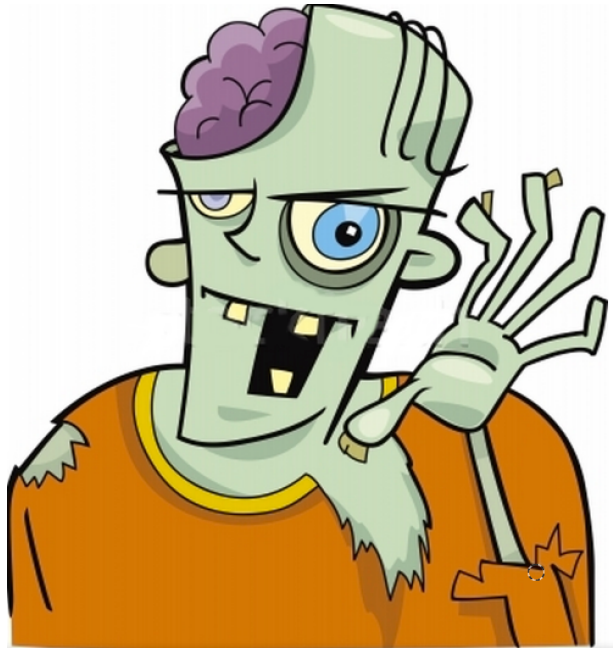
- **I** (Input feature map)
  - converts to bag-of-word-embeddings x.
- **G** (Generalization)
  - stores x in next available slot mN.
- **O** (Output) : Loops over all memories k=1 or 2 times:
  - 1st loop max: finds best match mi with x.
  - 2nd loop max: finds best match mJ with (x, mi).
  - The output o is represented with (x, mi, mJ).
- **R** (Response)
  - Ranks all words in the dictionary given o and returns best single word. (OR: use a full RNN here)



Memory Network



[Figure by Saina Sukhbaatar]



Memory Network

- **Representation of inputs and memories could use all kinds of encodings**
  - bag of words, RNN style reading at word or character level, etc.
- **Different possibilities for output module**
  - e.g. multi-class classifier or uses an RNN to output sentences.
- **If the memory is huge** (e.g. Wikipedia) we need to organize the m emories.
  - Hash the memories to store in buckets (topics).
  - Then, memory addressing and reading doesn't operate on *all* memories.
- **If the memory is full**, there could be a way of removing one it think s is most useless
  - **"forgets"** somehow.
  - That would require a scoring function of the utility of each memory..

Memory Network

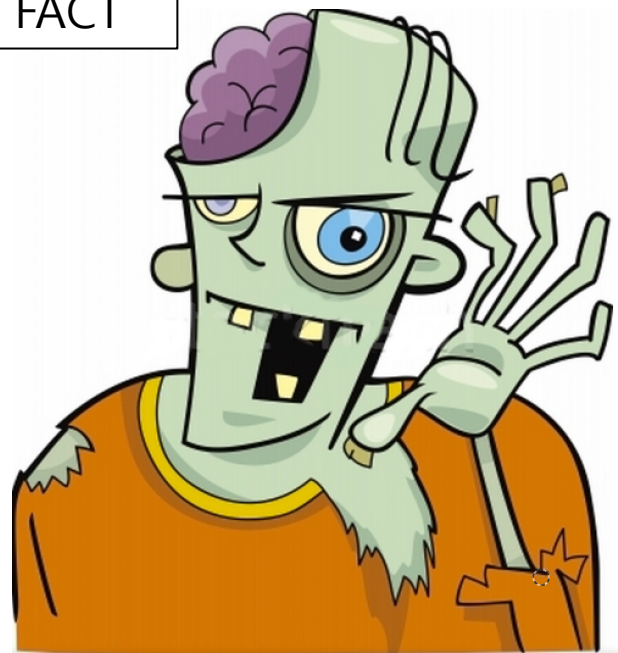
John is in the playground. ← SUPPORTING FACT

Bob is in the office.

John picked up the football. ← SUPPORTING FACT

Bob went to the kitchen.

Where is the football? A: playground



Memory Network

$$o_1 = O_1(x, \mathbf{m}) = \arg\max_{i=1,\ldots,N} s_O(x, \mathbf{m}_i)$$

$$o_2 = O_2(x, \mathbf{m}) = \arg\max_{i=1,\ldots,N} s_O([x, \mathbf{m}_{o_1}], \mathbf{m}_i)$$

$$r = \text{argmax}_{i=1,\ldots,|W|} s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], \bar{r}_i)$$

$$s_{O_t}(x, y, y') = \Phi_x(x)^\top U O_t{}^\top U O_t \left(\Phi_y(y) - \Phi_y(y') + \Phi_t(x, y, y') \right).$$

$$s(x, y) = \Phi_x(x)^\top$$

$$\sum_{\bar{f} \neq \mathbf{m}_{o_1}} \max(0, \gamma - s_O(x, \mathbf{m}_{o_1}) + s_O(x, \bar{f})) +$$

$$\sum_{\bar{f}' \neq \mathbf{m}_{o_2}} \max(0, \gamma - s_O([x, \mathbf{m}_{o_1}], \mathbf{m}_{o_2}]) + s_O([x, \mathbf{m}_{o_1}], \bar{f}'])) +$$

$$\sum_{\bar{r} \neq r} \max(0, \gamma - s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], r) + s_R([x, \mathbf{m}_{o_1}, \mathbf{m}_{o_2}], \bar{r}]))$$

Goal – Study – Plan

	Training				Implementation			
	week1	week2	week3	week4	week5	week6	week7	week8
Word embedding								
Sequence embedding								
Question Answering								
Memory Network								
Goal setting								
Dataset								
Implementation								