

APMA E4990 Project: PyFriend

A smart assistant to help you study Python

Team Members: Yiyi Zhang, Yilin Lyu, Shan Guan, Liangliang Huang

Background

In 2017, with 27% increase in views, Python stands alone as a tag that is both large and growing rapidly

When people have questions, most of them use Google and StackOverflow to find the answers

Google & StackOverflow

Google

- Output can be too complicated for beginners, not simple examples

StackOverflow

- Cannot recommend experts in similar topics
- Weak social network function

Data

	Id	OwnerUserId	CreationDate	Score	Title	Body
0	469	147.0	2008-08-02T15:11:16Z	21	How can I find the full path to a font from it...	<p>I am using the Photoshop's javascript API t...
1	502	147.0	2008-08-02T17:01:58Z	27	Get a preview JPEG of a PDF on Windows?	<p>I have a cross-platform (Python) applicatio...
2	535	154.0	2008-08-02T18:43:54Z	40	Continuous Integration System for a Python Cod...	<p>I'm starting work on a hobby project with a...
3	594	116.0	2008-08-03T01:15:08Z	25	cx_Oracle: How do I iterate over a result set?	<p>There are several ways to iterate over a re...
4	683	199.0	2008-08-03T13:19:16Z	28	Using 'in' to match an attribute of Python obj...	<p>I don't remember whether I was dreaming or ...

<https://www.kaggle.com/stackoverflow/pythonquestions/data>

Data (Tags)

	word	freq
3623	django	62818
11733	python-2.7	34616
10479	pandas	26854
11742	python-3.x	26814
9874	numpy	25848
8123	list	18951
8577	matplotlib	16521
12462	regex	14047
3495	dictionary	13413
15030	tkinter	10766



Project Goal

Part 1: Add tags to questions

Part 2: Help find the right questions for ambiguous input:

Match new questions to the existing questions, and return the top related existing Q & A

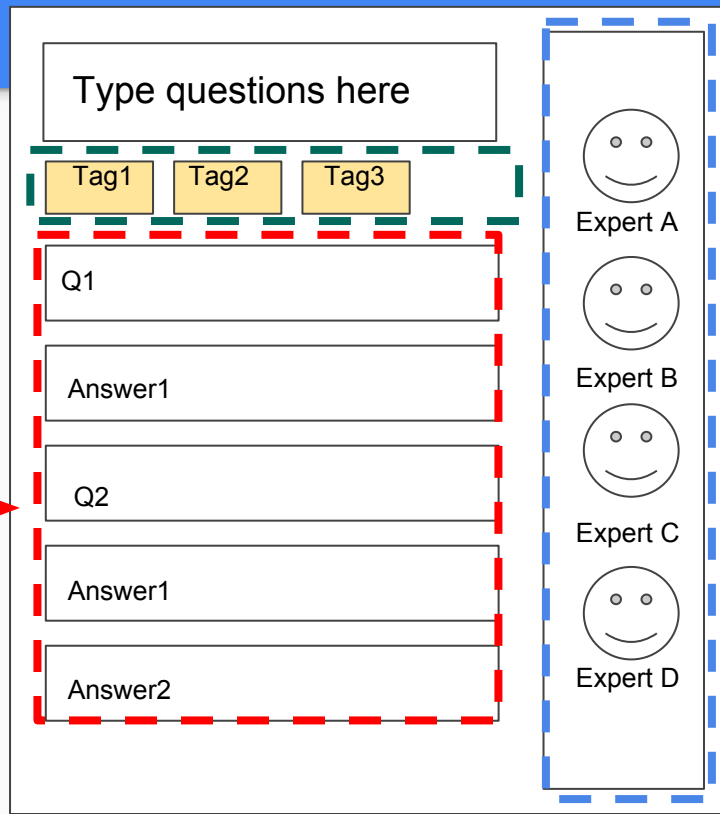
Part 3: Social Network function:

Recommend topic experts to the users

Website Snapshot

Auto-Tags of questions

Similar Questions Retrieval



Expert Recommendation

Auto-tag of query & retrieval: LDA topic model



Train a LDA model for auto tags for new queries



Match the queries with questions' tags(given in dataset)



Retrieve similar questions and answers to the questions



Explore tags and most FAQ of that tag

Content Based Recommendation Engine



Apply
TF-IDF for
the content
(both Q&A)



Calculate
similarities
(cosine)



Recommend
top experts in
the specific
areas (tags)

Further Improvement

1. Not only recommend experts to users, but also recommend questions to experts
2. Short text comprehension