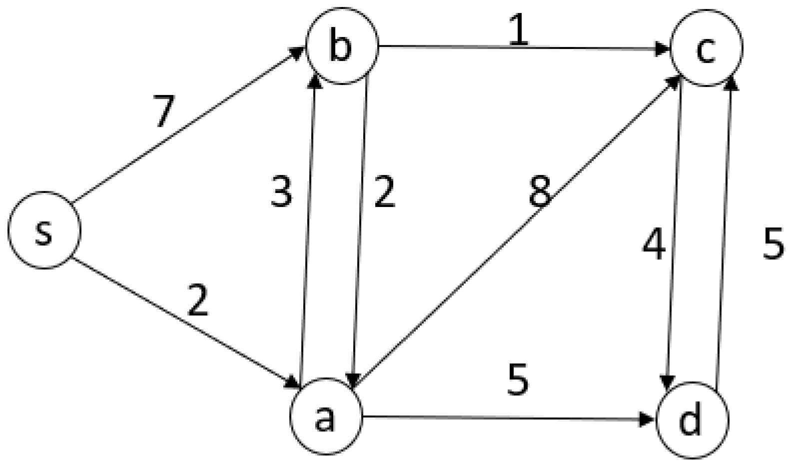
ÔN TẬP KHAI PHÁ WEB

I. BÀI TẬP:

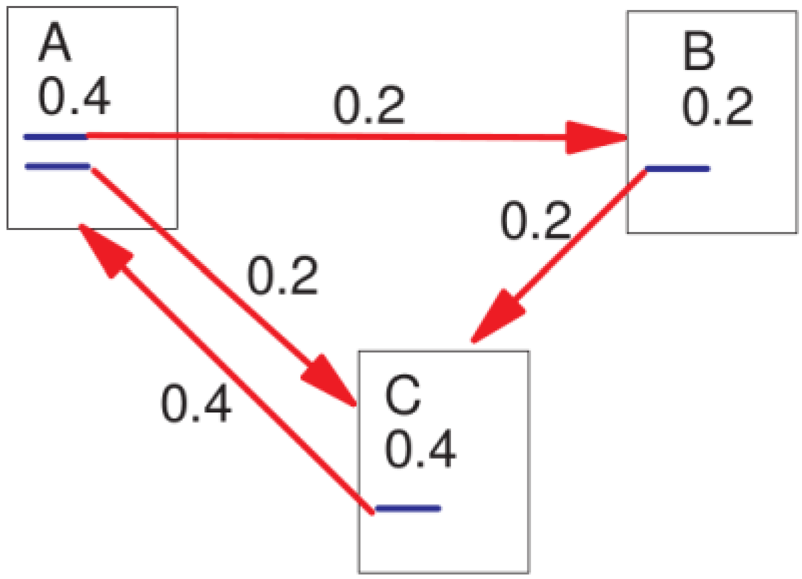
Bài tập 1:

Dijkstra: Khoảng cách ngắn nhất từ a tới c theo thuật toán dijkstra



Bài tập 2:

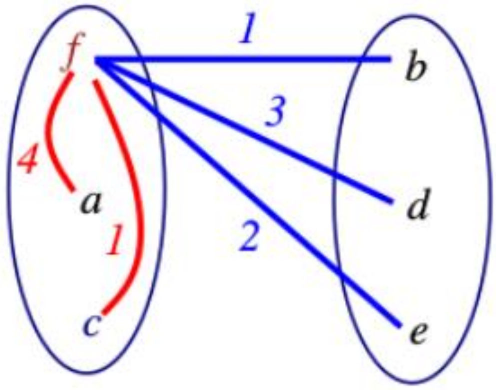
Pagerank: Giải theo hệ phương trình và theo phương pháp lặp (tới vòng lặp 4) với d = 0.8



Bài tập 3:

Kerninghan-Lin: Thực hiện một vòng lặp của thuật toán tìm lát cắt nhỏ nhất:

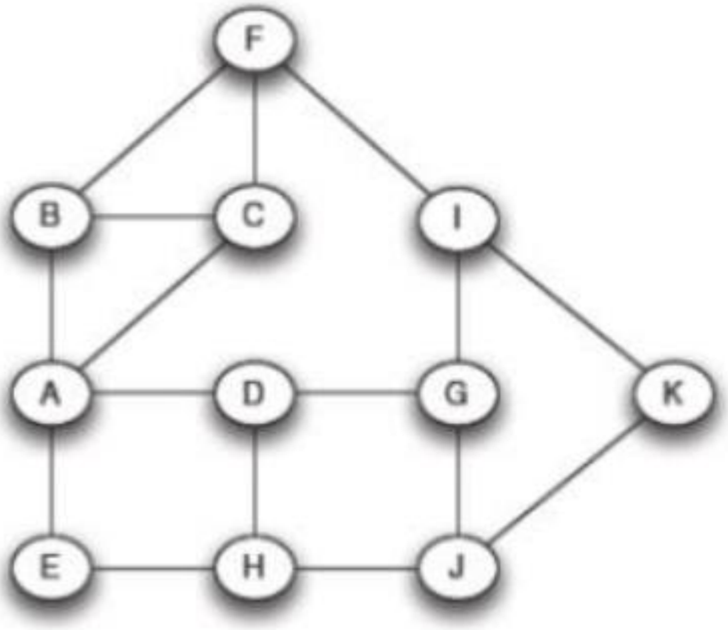
	a	b	c	d	e	f
a	0	1	2	3	2	4
b	1	0	1	4	2	1
c	2	1	0	3	2	1
d	3	4	3	0	4	3
e	2	2	2	4	0	2
f	4	1	1	3	2	0



Initial cut cost $= (1+3+2)+(1+3+2)+(1+3+2) = 18\ (22-4)$

Bài tập 4:

Tính khả năng thông qua của các cạnh dựa trên số đường đi ngắn nhất từ A tới các đỉnh còn lại trong đồ thị



Bài tập 5:

Tìm lớp của văn bản 5 dựa trên MNB với kĩ thuật làm mịn thêm 1

	docID	words in document	in $c = China$?
training set	1	Chinese Beijing Chinese	yes
	2	Chinese Chinese Shanghai	yes
	3	Chinese Macao	yes
	4	Tokyo Japan Chinese	no
test set	5	Chinese Chinese Chinese Tokyo Japan	?

Bài tập 6:

Tìm lớp của văn bản 5 dựa trên MNB với kĩ thuật làm mịn thêm 1

	docID	words in document	in $c = China$?
training set	1	Taipei Taiwan	yes
	2	Macao Taiwan Shanghai	yes
	3	Japan Sapporo	no
	4	Sapporo Osaka Taiwan	no
test set	5	Taiwan Taiwan Sapporo	?

Bài tập 7:

Dự đoán đánh giá của người dùng u_3 với sản phẩm i_4 theo phương pháp CF-knn dựa trên người dùng với k = 2. Độ tương đồng pearson, công thức tính dự đoán như trong slide bài giảng.

	i_1	i_2	i_3	i_4
u_1	5	4	4	1
u_2	2	1		
u_3	5	4	4	?
u_4		1	2	5

II. PAPER READING:

Hiểu rõ khái niệm, hiểu rõ các nội dung trình bày.

Paper 1:

In this paper, we present a feature-based named-entity recognition (NER) model that achieves the start-of-the-art accuracy for Vietnamese language. We combine word, word-shape features, PoS, chunk, Brown-cluster-based features, and word-embedding-based features in the Conditional Random Fields (CRF) model. We also explore the effects of word segmentation, PoS tagging, and chunking results of many popular Vietnamese NLP toolkits on the accuracy of the proposed featurebased NER model. Up to now, our work is the first work that systematically performs an extrinsic evaluation of basic Vietnamese NLP toolkits on the downstream NER task. Experimental results show that while automatically-generated word segmentation is useful, PoS and chunking information generated by Vietnamese NLP tools does not show their benefits for the proposed feature-based NER model

Paper 2:

In recent years, deep neural networks have yielded immense success on speech recognition, computer vision and natural language processing. However, the exploration of deep neural networks on recommender systems has received relatively less scrutiny. In this work, we strive to develop techniques based on neural networks to tackle the key problem in recommendation — collaborative filtering — on the basis of implicit feedback. Although some recent work has employed deep learning for recommendation, they primarily used it to model auxiliary information, such as textual descriptions of items and acoustic features of musics. When it comes to model the key factor in collaborative filtering — the interaction between user and item features, they still resorted to matrix factorization and applied an inner product on the latent features of users and items. By replacing the inner product with a neural architecture that can learn an arbitrary function from data, we present a general framework named NCF, short for Neural network-based Collaborative Filtering. NCF is generic and can express and generalize matrix factorization under its framework. To supercharge NCF modelling with non-linearities, we propose to leverage a multi-layer perceptron to learn the user–item interaction function. Extensive experiments on two real-world datasets show significant improvements of our proposed NCF framework over the state-of-the-art methods. Empirical evidence shows that using deeper layers of neural networks offers better recommendation performance.

Paper 3:

We apply recurrent neural networks (RNN) on a new domain, namely recommender systems. Real-life recommender systems often face the problem of having to base recommendations only on short session-based data (e.g. a small sportsware website) instead of long user histories (as in the case of Netflix). In this situation the frequently praised matrix factorization approaches are not accurate. This problem is usually overcome in practice by resorting to item-to-item recommendations, i.e. recommending similar items. We argue that by modeling the whole session, more accurate recommendations can be provided. We therefore propose an RNN-based approach for session-based recommendations. Our approach also considers practical aspects of the task and introduces several modifications to classic RNNs such as a ranking loss function that make it more viable for this specific problem. Experimental results on two data-sets show marked improvements over widely used approaches.