Topic-based Browsing of the Amazon Discussions Feedback Forum

Group Members

LAM Sing Ho, LIN Wing Ho

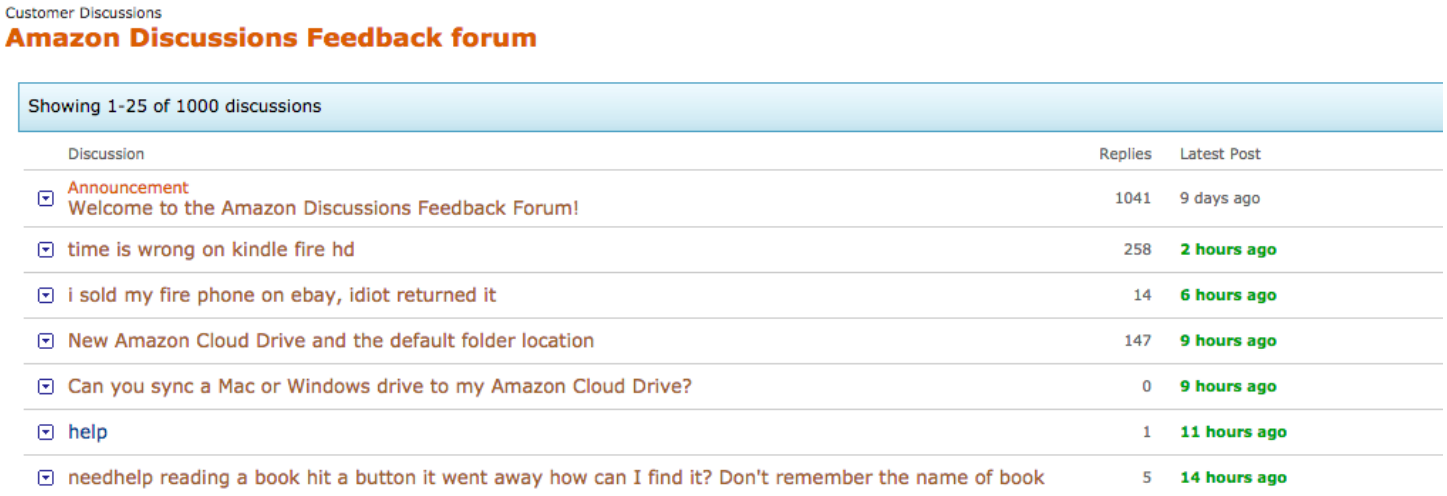
Supervisor

Professor Nevin L. Zhang, CSE, HKUST

Introduction

The ***Amazon Discussions Feedback Forum (Amazon Forum)*** is a platform for discussing matters relating to Amazon. With millions of registered users, this forum has three obvious problems:

1. Users can't quickly identify the **posts of interest**.
2. Users can't quickly identify the **current hot topics.**
3. Amazon can't place **effective advertisements** there.



Customer Discussions

Amazon Discussions Feedback forum

Showing 1-25 of 1000 discussions

Discussion	Replies	Latest Post
Announcement Welcome to the Amazon Discussions Feedback Forum!	1041	9 days ago
time is wrong on kindle fire hd	258	2 hours ago
i sold my fire phone on ebay, idiot returned it	14	6 hours ago
New Amazon Cloud Drive and the default folder location	147	9 hours ago
Can you sync a Mac or Windows drive to my Amazon Cloud Drive?	0	9 hours ago
help	1	11 hours ago
needhelp reading a book hit a button it went away how can I find it? Don't remember the name of book	5	14 hours ago

Figure 1: Amazon Discussions Feedback Forum

Objective

To solve the three problems, we developed a website called **"*Amazon Discussions Feedback Forum topic browser*"** so that users can browse this forum by topics, and receive relevant advertisements.

To do so, we:

- **Fetch** posts from the Amazon forum.
- **Detect** hidden topics from the posts using latent Dirichlet allocation (LDA), a topic model algorithm.
- **Link** each post to corresponding topics.
- **Display** the list of topics on a website.
- **Provide relevant advertisements** to each topic.

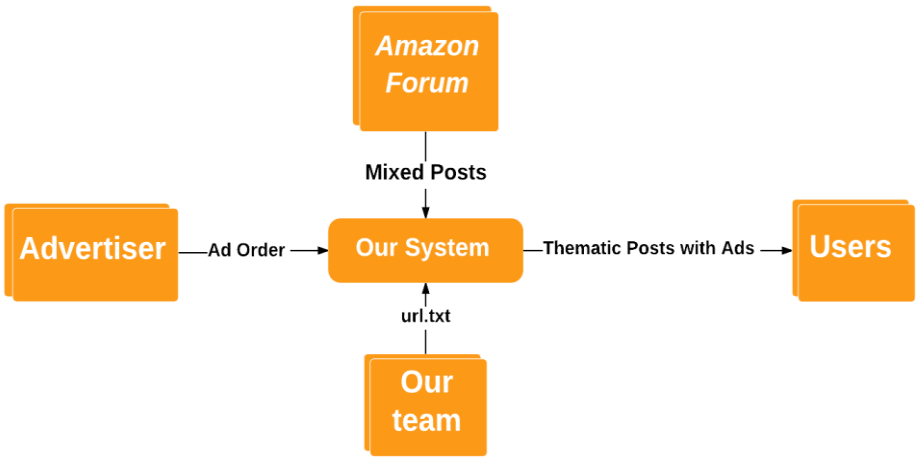


Figure 2: Showing the relationships of our system with relevant stakeholders

User interface

In the homepage, users can view the detected topics. They can then click on a topic to view the posts related to it.

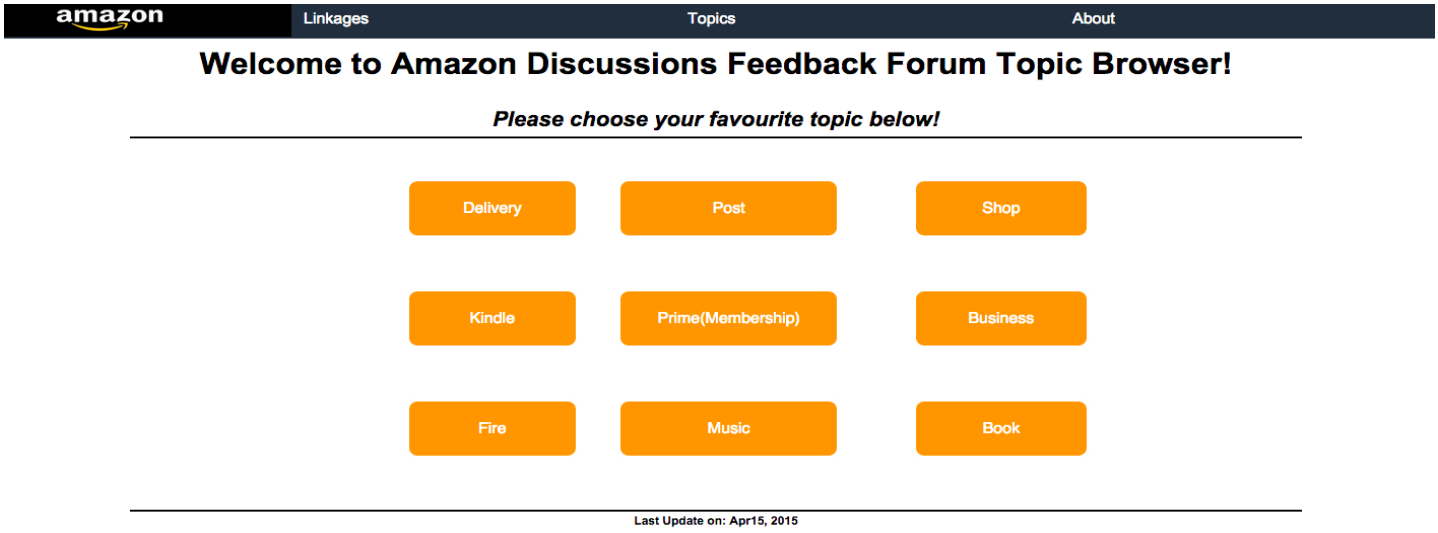


Figure 3: Homepage of our homepage

When users click on “Delivery”, the following webpage is shown. The middle column shows the posts related to “Delivery”. A **navigation bar** (left) and a relevant **advertisement** (right) are also displayed.

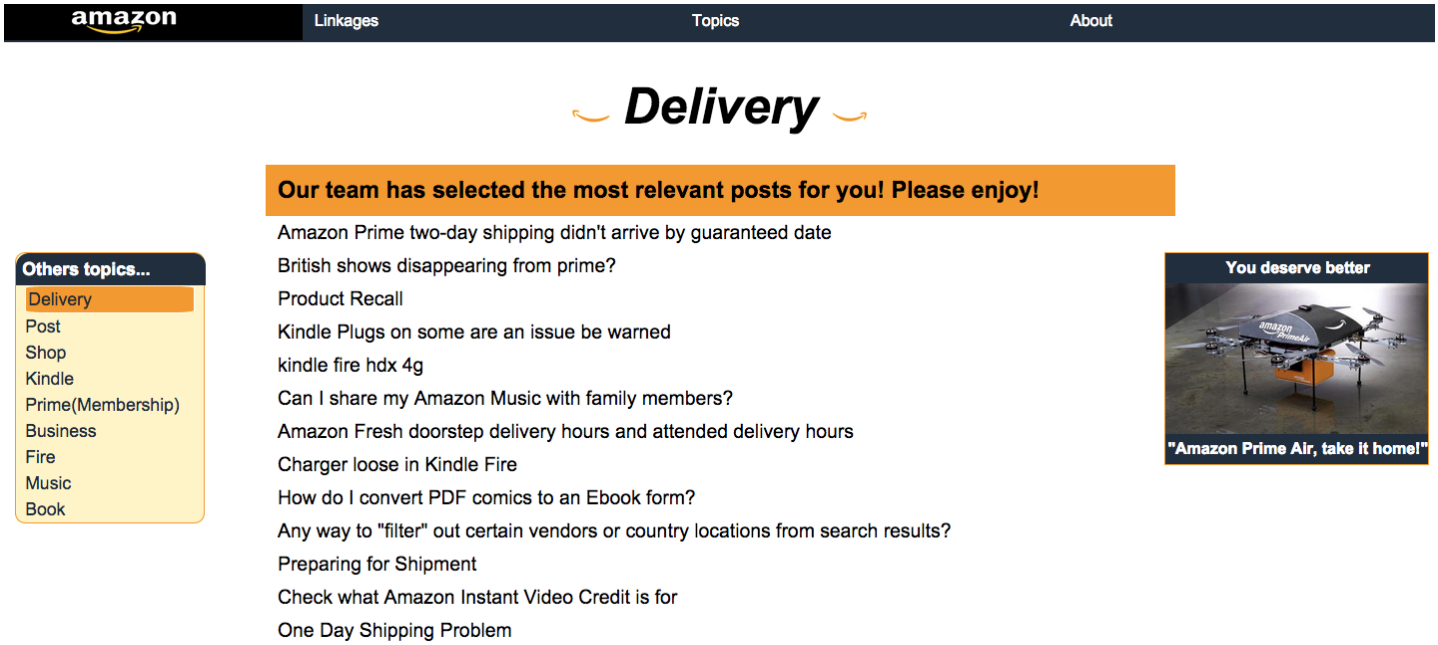


Figure 4: “Delivery” page with a related ad

Principle

To **automatically detect hidden topics** from the posts, we used a well-known probabilistic topic modeling algorithm – latent Dirichlet allocation (**LDA**), which is particularly suitable for processing text data.

In LDA, we let $\{\beta_1, \beta_2, \dots, \beta_k\}$ be the hidden topics, and assume that each word is drawn from the identical vocabulary set $\{1, \dots, V\}$. A document is represented by a sequence of words $\{W_1, W_2, \dots, W_n\}$. Each word is generated by one topic $z = \{1, \dots, k\}$. A Dirichlet prior $Dir(\alpha)$ is introduced to the K-dimensional document-topic distribution $\theta^{(d)}$. The following is the generative process for a N-word document d:

- Choose topic proportions $\theta \sim Dir(\alpha)$
- For each word W_n:
    1. Choose a topic assignment $z_n \sim Mult(\theta)$
    2. Choose a word W_n from $p(W_n | z_n, \beta)$

The relationship among different parameters are illustrated below:

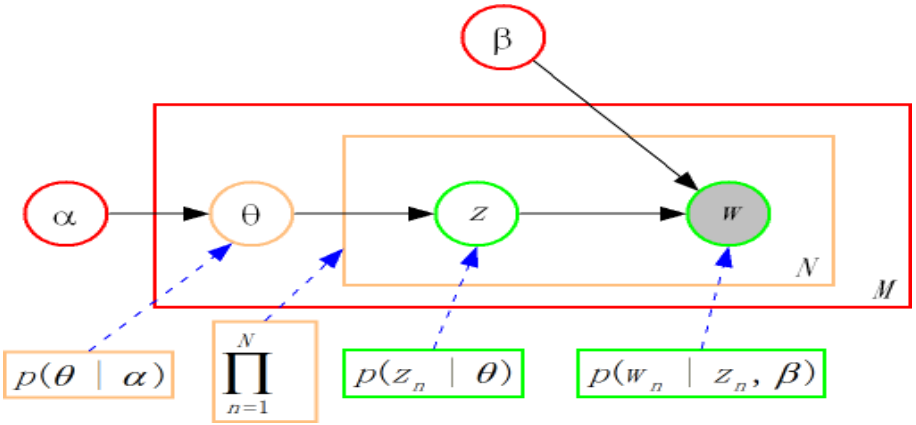


Figure 5: Pictorial view of LDA

Summary

In this project we developed a topic-based browser for the forum, with an aim to help save the users' precious time.

We have:

- **Fetched** 500 posts with 4394 replies.
- **Detected** 9 thematic topics.
- **Established** the linkages between the topics and the relevant posts.
- **Displayed** the posts in a **user-friendly** website.
- **Placed relevant advertisements** with a personalized message.

Fetching

Detecting

Linking

Displaying

Advertising