

Some Problems in Web, Email, and Literature Domain

The order of the problems, also serve as the order of my preferences.

1. Web Domain:

Name: Automatic News and Information Extractor, Classifier, and Comparator.

User: Ordinary Internet user, who is looking for a way to read news in a more classified and organized way from all the sources he/she wants, spending the minimum amount of time.

Data Involved: The specified web pages by the user, input data entered by the user.

Function: This problem was motivated by the problem that I am having almost every day. There are two domains in which I check news. First social and political news, and second sports news. For each of the categories above, I check several online sources to get the desired breadth and depth of information. However, this process takes a lot of my time everyday. What I am looking for, is a software (or a web page) that gets the links to the online news/information sources that I use once, and do the following actions either on a regular basis, or in an on demand fashion:

1. Regular (or daily) function: Extract the mutually related news from all the sources and provide me with titles, sources, summaries, and links to the full content/articles. It is expected that the software do this separately for every interesting, hot, or commonly discussed issue that appears in the media.
2. On demand function: It should provide the user with the ability to look for all the material related to a specific issue/topic that user enters. It should somehow extract (possibly from archive), reformat, and summarize all the related material to the issue specified by the user (only in the specified sources).

Result: As a result, when I get up in the morning, instead of opening 7-8 different pages everyday, I just open a single page, which already knows the sources that I am interested in. It extracts, reformats, and summarizes everything for me, and shows them to me including the reference. It can also, as a special link, extract and show information that might only occur in one or some of the specified sources. It can also recommend some additional links to the user.

2. Email Domain:

Name: Intelligent Email Organizer

User: Any faculty, poor TA, or business person who is very busy and is always dealing with several tasks, and receives many emails every hour!

Data Involved: The emails received, and also the input data (can include people's schedule)

Function: This problem was motivated by another problem that I have every day. I look at the problem as a Grad student who is taking 3 courses and is TAing a class with many students, and at the same time he is doing some research with some other faculty. At the same time he has many tasks, and receives many emails related to each of the tasks in different times of the day. This student decides to divide his hours in the week into spots for each of his responsibilities (probably 2-3 spots for each responsibility). Therefore, although he wants to be accessible by all the people who need him at any time, he wants to only focus on what he has in his schedule, and not be distracted by other tasks, and emails.

So, what the system does is this: It gets the user's schedule, the areas in which the user wants its tasks to be categorized, the relation between the areas and the available spots, and some other clues. Then after receiving emails from different people, it categorizes them and shows them to the user only at the time that the user has allocated for those kind of tasks. It can also extract meetings, and alert the user...and do many many (I really mean it!) interesting, and time preserving tasks!

3. Literature Domain:

Name: Starting Point for Research in Any Area

User: Any faculty, or student who is looking at entering a new are of research.

Data Involved: All the papers/or indexed summaries available on the web.

Function: Whenever a researcher wants to enter a new area, he/she faces a big question: How or from where should I start? Finding an answer for this question can be a difficult or at least time consuming task. It would be great if a system exists that can gather, and summarize all the information about all the papers (including classic, highly referenced, cutting edge, etc) and also all the people that work in the related areas (including summaries and information about their publications, projects, affiliations, etc). This intelligent system can somehow generate a route through which the user can get all the information he/she would need in order to start getting into the desired area.