Problem Giving Articles to User

Description of the problem ..

System starts up and processes RSS feeds

* That results in about 2000 Articles being stored
* The first process rss feeds takes a long time - I don’t have exact number but maybe 10 minutes.
* **Subsequent process rss feeds only take 20 seconds** - I’ve checked this on the UCD VMs - there is scheduler log entry for start/finish
* So I don’t think handling celery jobs better will be that useful given the short time left in the project

User logs in and selects topics

* This calls a function (all in web container) to get article ids matching the topic and store in UserArticles (link to Article is added not whole Article).
* Note the get article ids we currently limit to 10 (this is configurable though on Django Admin via SystemConfigs)
* Storing a small number makes a lot of sense because it takes a long time to store or delete lots of UserArticles. EG 1000 for example.
* However only getting 10 article ids is not great - because we never consider the other articles. A better way would be to get all the article ids (no time penalty as its redis), rank them all, then store the top 10 ranking. This method may not take a lot of time as no extra DB hits. Probably there is other ways to achieve this. The point about this is that we want to consider all articles and not just give the user first 10.

Processing new RSS feeds after first one

* For new articles, process rss feed will match for every user and add articles to UserArticles. It’s not a lot of articles. Process RSS feeds doesn’t consider existing articles.

Problem:

* The problem is that I login, select topics, and get my 10 articles. Note again that 10 is configurable - I will check at work what number people would prefer to see. (if you look at SystemConfig in the Admin there is separate config for number stored and number of articles shown to user).
* I read articles, great - newfast is the best thing ever
* I come back tomorrow expecting great stuff, and its the same 10 articles with maybe 1 or 2 different. The only different articles will be new articles added by process rss feeds. So its a bit of a rubbish user experience - articles not changing much.

Possible changes - there are a number of things we can do. I think if we think of a clever solution .. it won’t be much work and will improve the system a lot. At this stage we need to think purely about the demo I think. Some thoughts here:

* Give the user the real 10 highest ranking by considering all articles
* Assuming we display 10 and store 50 (again remember these number are configurable).
* Downvote articles we display for the User - so each page reload would see some of the 10 displayed dropping down and some of the 40 moving up.This should keep the news page fresher by giving new articles.
* For demo the above might be enough .. you might also think what to do when the 50 are all displayed
* One nice feature - easy to do - and configurable would be: when displaying articles, to make every 5th article one that doesn’t match their topics - so would be taken from highest ranking articles not in UserArticles. That would keep it interesting I think.