**Team C+D – progress report and wishlist**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Aim** | **Task** | **Person** | **Results/issues** | **Next steps** |
| Run search across full data set to find adverts containing a key word | Investigate best way to identify adverts within OCR | David/Alex/Olivia | Looked for “blanks” in OCR to indicate areas on pages with design/borders/non-standard features. Some success, but seems inconsistent | Investigate other ways to identify adverts?  Meanwhile carry out searches across both adverts and articles |
|  | Run search across first 5000 issues for “asthma” | David | Completed, but results show gap 1895-1924.  Gap is because issues are not in chronological order – are 1895-1924 the 5001-6000+ issues? | Run second search to fill the gap |
|  | Run search on 5000+ issues for “asthma” | David | Produced results from 1895-1903, but still gap 1904-1925 | Figure out why there’s still a gap |
|  | Investigate Elasticsearch as a more effective means of retrieving results | David | Issue with uploading full-text from each issue is that there’s no way of knowing what page a search hit occurs on. Issue with Bonsai cloud service is that we uploaded too much stuff and got shut down. | Upload each page of each issue individually for a more useful index. |
| Run search that produces a smaller set of results (related to asthma) to allow work on less cluttered visualisation | Run search across full data set to find “inhaler” | David | Same gap within initial search (1895-1924). Second search fills this |  |
| Create timeline visualisation from search results | Use results from “asthma” keyword search to create timeline. | Olivia | Successful, although sheer quantity of results presents issues.  Timeline overviews all results – individual images can be inspected by zooming in. | Create wishlist of filters/sub searches that would thin out the results |
|  | Refine existing timeline to improve accessibility/interpretation | Olivia | * Years and gridlines added. * Results organised according to date/issue within each year. * C+D is a regular publication so patterns and trends are made visible. (Runs of 6 adverts, advert campaigns for a brand in winter, C+D shifting from published monthly to weekly, changing paper tone over time). * Date of each image added as a tooltip | * Find a way to separate out multiple results from same issue (currently “stacked” on top of each other, so invisible) * Find a way to filter results by either adverts or articles * Introduce highlighted search term within each page image * Find a way to “cluster” the same/similar adverts together 1) for comparative research 2) to de-clutter the timeline * Add links to the pages in universal viewer |
|  | Use results from “inhaler” keyword search to create timeline | Olivia | Successful and fewer results makes the visualisation less cluttered.  Years, gridlines and month names added  Stripes added to distinguish each month | * Use this sparser data set to try further searches e.g. 1) "asthma" 2) an illustration featuring "asthma" 3) brand name "Ventolin". |
|  | Mock up further searches and/or filters of “inhaler” timeline to explore visualisation | Olivia/Alex | Further searches run on small sections of “inhaler” results to create subsets: 1) "asthma" 2) an illustration featuring "asthma" 3) brand name "Ventolin". | * More work on mock ups * Work out why the “Ventolin” search is unsuccessful |
| Consider interpretation for timeline | Create timeline of milestone drug development for asthma treatment that could act as “fixed points” on timeline | Briony | Done - taken mainly from RPS Museum “Developing Treatments” document | Look for a key representative image (search result from C+D) that could accompany the fixed point visually |
|  | Find key images for asthma milestones | Briony | * Completed, although issues around finding something visual that is in the year of drug launch to illustrate the milestones most effectively * Olivia and Briony discussed approaches to including interpretation in timelines | * Possible use as “fixed” image in visualisation to aid interpretation or to be highlighted in some way to provide some (optional?) structure to timeline * Could be used to accompany (optional?) interpretation – see row below. |
|  | Draft potential introductory/explanatory text that could accompany the timeline | Briony | Draft completed, but dependent on other aspects of visualisation | * Think further about appearance/purpose/usage of visualisation. * Further discussions around nature of interpretation vs. uninterpreted search results |
|  | Refine overarching research question into sub-questions | Briony | Some examples created with ideas of searches and related issues – see separate *Sub research questions* document | Think further about how these (and other?) sub-questions link to potential filters/themes in visualisation. Link with work on searches/filters above. |

**Wishlist**

1. Time and resources for a back end developer to work on search functionality. The basic issue is that Chemist and Druggist is currently treated as one volume (b number) which means that every search attempt is looking through 500,000+ pages