



## Programme

### EBI Workshop

6th October 2014

|                 |                                                                                                                                                                                                  |
|-----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 9:30 - 10:00am  | <b>Introduction</b> <ul style="list-style-type: none"><li>- Welcome (JCM, SK)</li><li>- Introduction to ContentMine (PMR)</li><li>- Delegates' introductions and suggested use cases</li></ul>   |
| 10:00 - 11:00am | <b>Ice-breaker - manual markup exercise (PMR)</b> <ul style="list-style-type: none"><li>- Split into pairs or small groups and annotate a paper</li><li>- Discussion</li></ul>                   |
| 11:00 - 11:15am | <b>Tea/coffee break</b>                                                                                                                                                                          |
| 11:15 - 11:45am | <b>Talk - architecture of content mining (PMR)</b>                                                                                                                                               |
| 11:45 - 12:15pm | <b>Talk - legal aspects of content mining (CO)</b>                                                                                                                                               |
| 12:15 - 1:15pm  | <b>Lunch break</b>                                                                                                                                                                               |
| 1:15 - 1:45pm   | <b>Demonstration - entity extraction in chemistry</b> <ul style="list-style-type: none"><li>- simple regexes</li><li>- phylogenetic trees</li><li>- chemistry</li><li>- hands on demos</li></ul> |
| 1:45 - 2:15pm   | <b>Demonstration - document section tagger</b> <ul style="list-style-type: none"><li>- how to turn semistructured documents into normalised XML</li><li>- the role of SectionTagger</li></ul>    |
| 2:15 - 2:30pm   | <b>Tea/coffee break</b>                                                                                                                                                                          |
| 2:30 - 3:00pm   | <b>Demonstration - scrapers</b> <ul style="list-style-type: none"><li>- quickscrape</li><li>- running, modifying, and building a scraper</li><li>- hands on demo</li></ul>                       |
| 3:00 - 4:00pm   | <b>Group hacking</b>                                                                                                                                                                             |
| 4:00 - 5:00pm   | <b>Discussion and strategy</b>                                                                                                                                                                   |