Text and Data Mining

Reed Elsevier perspectives

Who is doing TDM on content published by Elsevier?

Text mining licences with:

Higher education/research institutions

Corporates

App developers
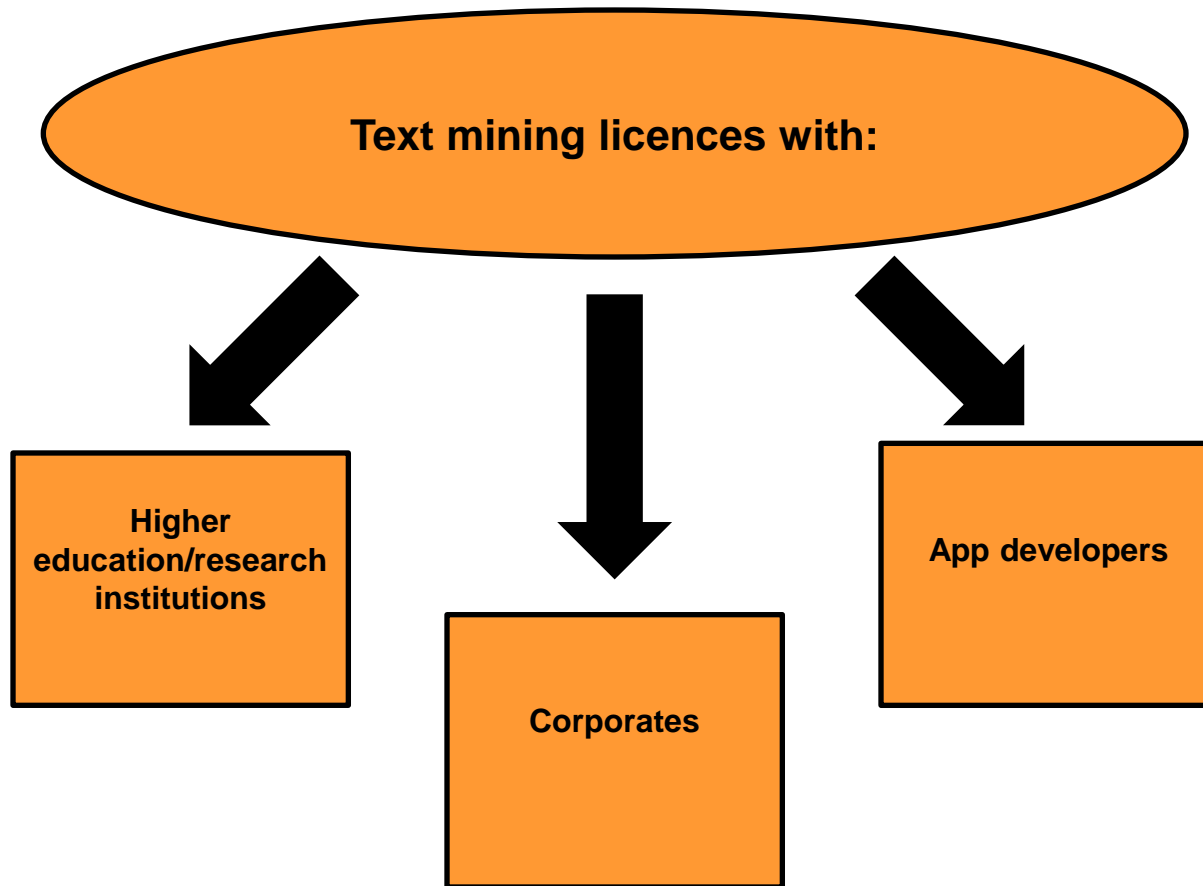
How do we give text miners access?

Transfer of content to user's system enabled by one of two delivery mechanisms under a separate licence:

Via API

ConSyn
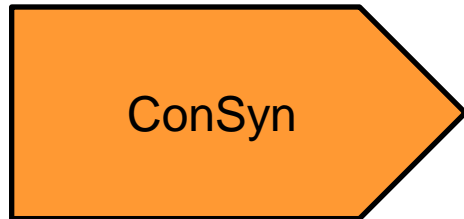
- Delivery mechanisms have set up and running costs
- To recoup costs our policy to date has been to charge the value add of the services to commercial users

Why is content delivery managed this way? (1)

Platform Stability

- Miners want content at scale – they generally don't do TDM on a couple of articles

- Delivering scale via our main platform ScienceDirect.com is suboptimal

- APIs or ConSyn the solution – leave ScienceDirect untouched

- Effectively we are separating machine to machine traffic from traffic created by real users going to ScienceDirect.com

Content Security

- Free for all on ScienceDirect would not allow bona fides to be checked

- XML versions less susceptible to piracy than PDFs

Why is content delivery managed this way? (2)

It's more efficient for genuine text miners

- Most miners prefer to work off XML, not from article versions on ScienceDirect

- Our delivery mechanisms put the content into their hands fast

- Scraping 1000s of articles suboptimal for miners too

TDM Outputs

- CC-BY-NC license on output when redistributing results of text mining in a research support or other non-commercial tool

- Require DOI link back to mined article whenever feasible when displaying extracted content

- Grant permission to use snippets around extracted entities to allow context when presenting results, up to a maximum of 200 characters or 1 complete sentence

Next steps – encouraging and simplifying text mining

- **Continue to enable mining for users at subscribed institutions where their purpose is non-commercial without charge.**

- **Actively encourage text mining requests from researchers and institutions to our single point of contact - TDMSupport@elsevier.com**

- **Set up a pilot automated licensing system**

Summary

Licensing working well at Elsevier - will improve further

- Demand to mine being met, no extra charges in the vast majority of cases
- Additional services to support mining likely to be offered as we improve understanding of miners' needs

It's early days

- Mining demand is embryonic – low numbers at the moment

Copyright exceptions, a big cause for concern

- Major risk of spike in unauthorised redistribution.
- Platform stability threatened
- Risk of chilling effect on future service innovation

Thank you

alistair.tebbit@reedelsevier.com