manipal ProLearn

Text Analytics

Introduction to Topic Modelling

By

Kathirmani Sukumar

Definition

*"Topic modelling is a type of statistical model for discovering the abstract **topics** that occur in a collection of documents" - Wikipedia*

"Topic modelling helps to discover and annotate large documents with hidden topics" – Lin Liu et al

Example

- Amazon has great customer support. Very quick response — Customer Support
- Delivery speed is good. Great packing — Delivery
- Got a damaged packing from Amazon. Need a replacement — Delivery
- Customer support promptly responded back to my queries — Customer Support
- Amazon prime offers quick delivery with free of cost — Delivery
- Deals are quite good. Good offers — Deals

Terms define topics

Customer Support

- Support
- Service
- Call center
- Response
- Polite

Delivery

- Speed
- Packing
- Damaged
- Protection

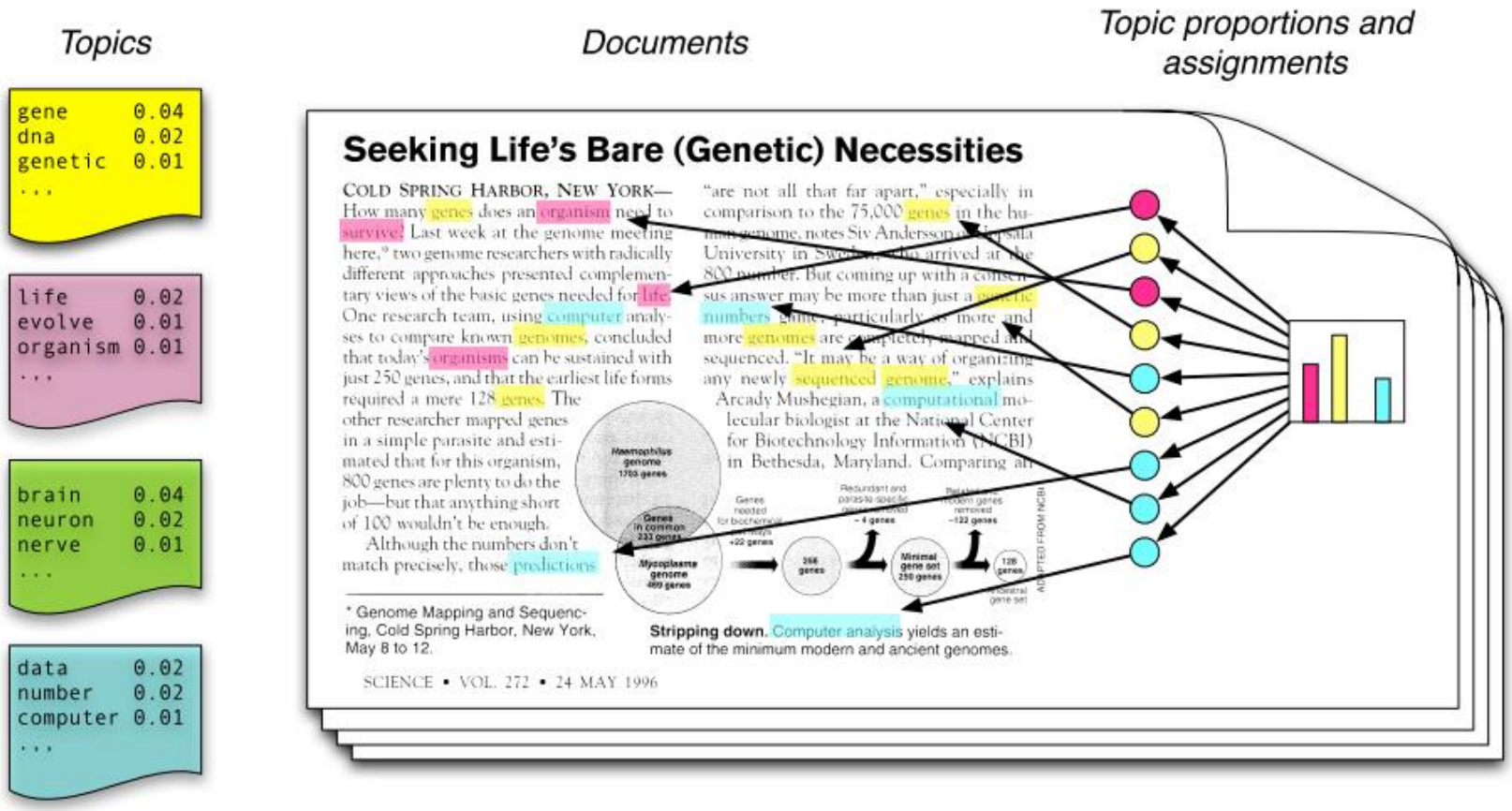
Deals

- Offers
- Sales
- Promotions
- Coupons
- Offer

Amazon reviews

- The first group contains words related to tablet (ipad, android, app, device etc). Hence we can label it as ***Tablets***
- The second group contains words related to e-reader (nook, kindly, books, touch, page etc). Hence we can label it as ***E-Reader***
- The third group contains words related to battery problems. We can name it as "***Batteries***"
- The fourth group contain words related to wall mounting. We can name it as "**Wall mounting**"

Latent Topics



Popular Methods

- Latent Sematic Analysis (LSA)
- Latent Dirichlet Allocation (LDA)

Applications

- Clustering
  - Cluster documents in to different groups
  - Creating catalogue automatically for products, books based on description and content
- Multi-Tagging
  - Tag files, emails, SMS automatically
- Reduce Dimension
  - Group words together to reduce dimension in Document Term Matrix

manipal PROlearn

THANK YOU