XTractor

Data mining simplified

William G Town

ICIC Nice October 2008

Molecular Connections Pvt. Ltd.,

2/2, Kandala Mansion, Kariappa Road,
Basavanagudi,
Bangalore - 560 004, India
phone :- +91 80- 4120 5016
fax No :- +91 80- 4120 8956
http://www.molecularconnections.com

For Free Access visit:
www.xtractor.in

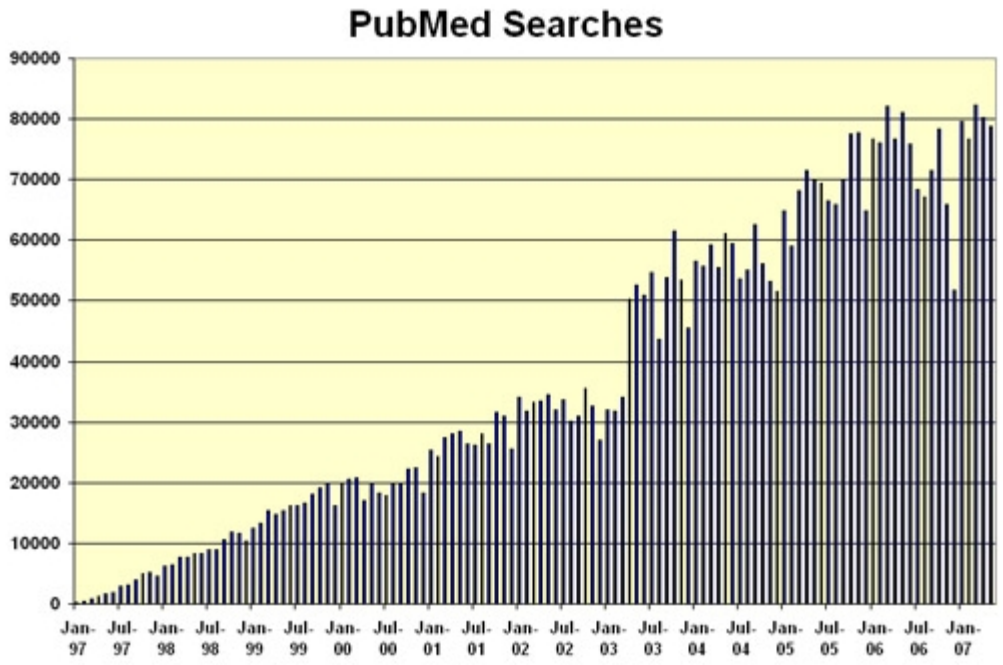
Problem statement

1. Mining PubMed data regularly is time consuming and expensive

1. Most of the hand curated databases are not current (provide updates once in 3-6 months) and are expensive

1. Text-mining technologies are relatively less accurate with some false positives

PubMed



(Graph as published by NCBI)

PubMed searches have increased dramatically in the last 10 years - an average of 70,000 searches every month. About 12,000 articles are added every week.

For Free Access visit:
www.xtractor.in

Introduction

- **XTractor™** is the first of its kind – a regular PubMed alert service which also provides manually curated and annotated sentences for user selected keywords

- **XTractor™** enhances scientists' experience of reading literature and aids in:

  - **Abstract summarization**
  - **Categorization**
  - **Topic tracking**
  - **Concept linking**
  - **Customization**
  - **Report generation**
  - **Information visualization**

For Free Access visit:
www.xtractor.in

Manual mining - conventional approach

- Search PubMed regularly ↓
- Download the hits obtained ↓
- Manually scan OR use text-mining for relevancy of papers

For Free Access visit:
www.xtractor.in

Manual mining - conventional approach

2: Breast Cancer Res Treat. 2008 Apr 18 [Epub ahead of print] Related Articles

SpringerLink FULL-TEXT ARTICLE

Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis.

Lu WL, Jansen L, Post WJ, Bonnema J, Van de Velde JC, De Bock GH.

Department of Epidemiology, University Medical Center Groningen, University of Groningen, PO Box 30001, 9700 RB, Groningen, The Netherlands.

Purpose The purpose was to establi... a local recurrence of breast cancer as compared to late detection. Design A meta-analy... nager software (RevMan version 4.2). Studies were included if women were treated for ... ant metastasis at primary diagnosis and if these concerned routine follow-up strateg... ecurrences. Data regarding the risk for death were derived from each study. Multi level models were used to study heterogeneity by using MLWin. Results Thirteen studies concerning 2,263 patients were included. Early detection of breast cancer recurrences during follow-up gave a significantly better survival as compared to late detected recurrences (HR: 1.68 (95% CI: 1.48-1.91)). Survival was better when the recurrence was found by mammography instead of physical examination or in patients without symptoms as compared to those with symptoms (HR: 2.44 (95% CI: 1.78-3.35); HR: 1.56 (95% CI: 1.36-1.79), respectively). If all breast cancer recurrences would be detected earlier, that 5-8 deaths (i.e. an absolute reduction in mortality of 17-28%) would be avoided by performing routine follow-up during a 10 year-period for 1,000 breast cancer patients. Conclusion These data support the hypothesis that detection of isolated loco-regional or contra-lateral breast cancer recurrences in patients without symptoms has beneficial impact on survival of breast cancer patients when compared to late symptomatic detection.

PMID: 18421576 [PubMed - as supplied by publisher]

Step 2

- Analyze relevant abstracts
- Summarize the findings
- Extract required facts

For Free Access visit:
www.xtractor.in

Manual mining - conventional approach

2: Breast Cancer Res Treat. 2008 Apr 18 [Epub ahead of print] Related Articles

SpringerLink FULL-TEXT ARTICLE

2: Breast Cancer Res Treat. 2008 Apr 18 [Epub ahead of print] Related Articles

SpringerLink

2: Breast Cancer Res Treat. 2008 Apr 18 [Epub ahead of print] Related Articles

SpringerLink FULL-TEXT ARTICLE

Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis.

Lu WL, Jansen L, Post WJ, B

Step 3

Department of Epidemiology, Uri of Groningen, PO Box 30001, 9700 RB, Groningen, The Netherlands.

Purpose The purpose was to establish the impact on survival of early detection of a local recurrence of breast cancer as compared to late detection. Design A meta-analysis was carried out using Cochrane review manager software (RevMan version 4.2). Studies were included if women were treated for primar distant metastasis at primary diagnosis and if these concerned routine follow-up strategies urrences. Data regarding the risk for death were derived from each study. Multi level mo ng MLWin. Results Thirteen studies concerning 2,263 patients were included. Early det llow-up gave a significantly better survival as compared to late detected recurrences as better when the recurrence was found by mammography instead of physical examina compared to those with symptoms (HR: 2.44 (95% CI: 1.78-3.35); HR: 1.56 (95% CI: 1.36-1.79), respectively). If all breast cancer recurrences would be detected earlier, that 5-8 deaths (i.e. an absolute reduction in mortality of 17-28%) would be avoided by performing routine follow-up during a 10 year-period for 1,000 breast cancer patients. Conclusion These data support the hypothesis that detection of isolated loco-regional or contra-lateral breast cancer recurrences in patients without symptoms has beneficial impact on survival of breast cancer patients when compared to late symptomatic detection.

PMID: 18421576 [PubMed - as supplied by publisher]

Public Databases

Create Databases

Public Databases

- Store the extracted contents in a database
- Map to ontologies
- Retrieve and analyze

For Free Access visit:
www.xtractor.in

Time costs

Keywords searched across PUBMED	Dates of addition in PUBMED	Number of abstracts	Approx time taken for manual annotation and extraction (10 minutes per abstract)
Breast neoplasm	*Last 90 days*	*621*	*6210 mins OR 103 hours OR 11 working days*
	Last 60 days	*271*	*2710 mins OR 45 hours OR 5 working days*
	Last 30 days	*56*	*560 mins OR 9 hours OR 1 working day*

It takes at least one working day to extract all the possible relations from 56 abstracts

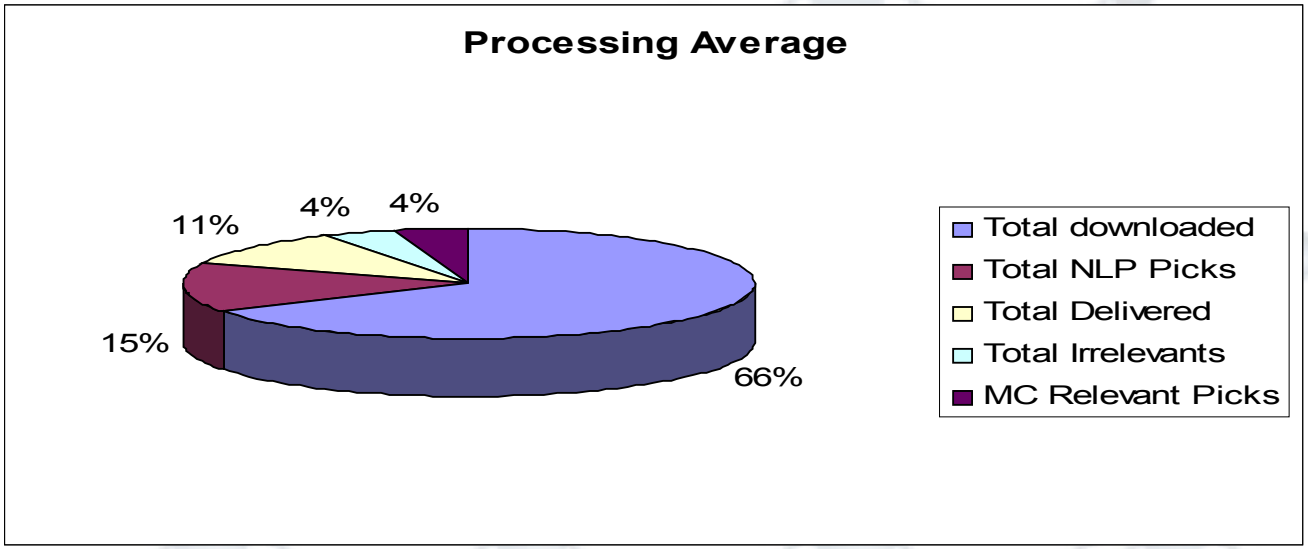
Analysis conducted as on 28 Apr 08 on PubMed

Case Study:

To clear 30 days of data on breast neoplasm, it would take at least 1 man day (based on our past estimates that it takes at least 10 -12 minutes to annotate one abstract).

For Free Access visit:
www.xtractor.in

Xtractor vs NLP



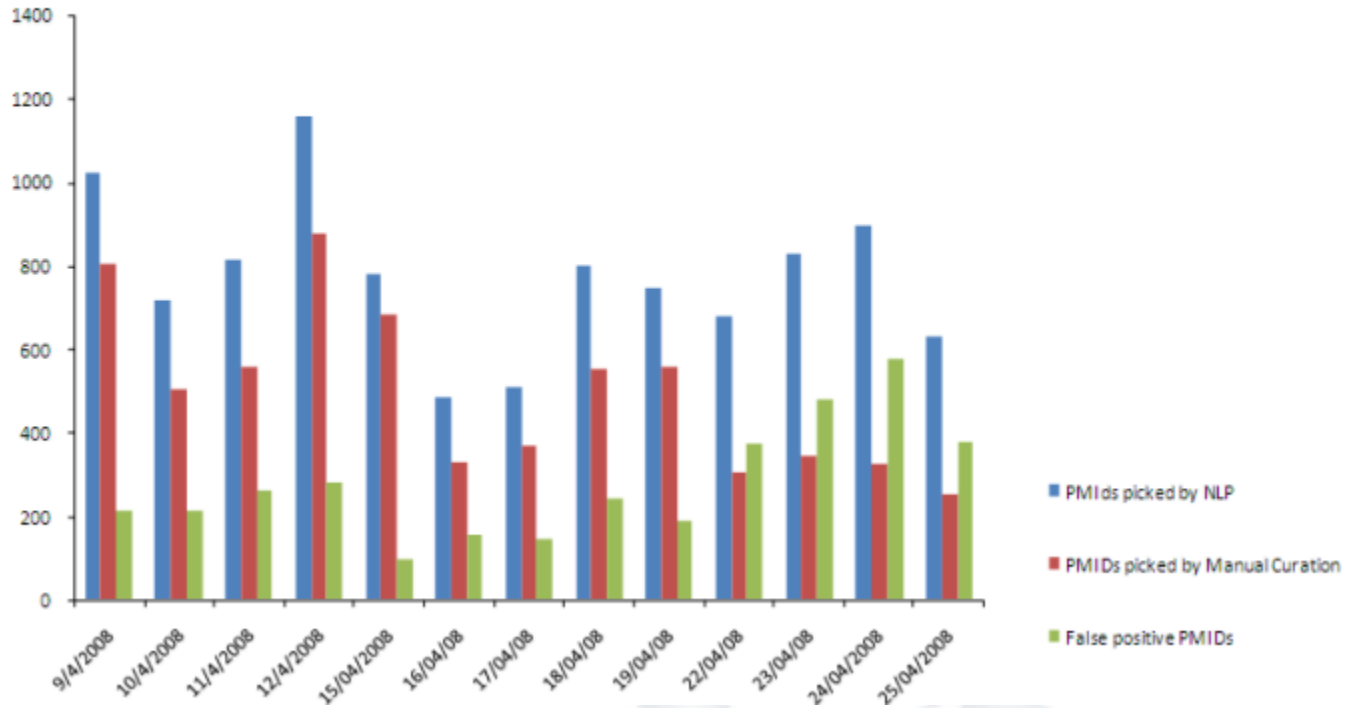
	Avg/Day
Total Abstracts Downloaded	4090
Total NLP picks	917
MC relevant picks	250

A case study on daily disease and protein annotation

For Free Access visit:
www.xtractor.in

XTractor better than NLPs

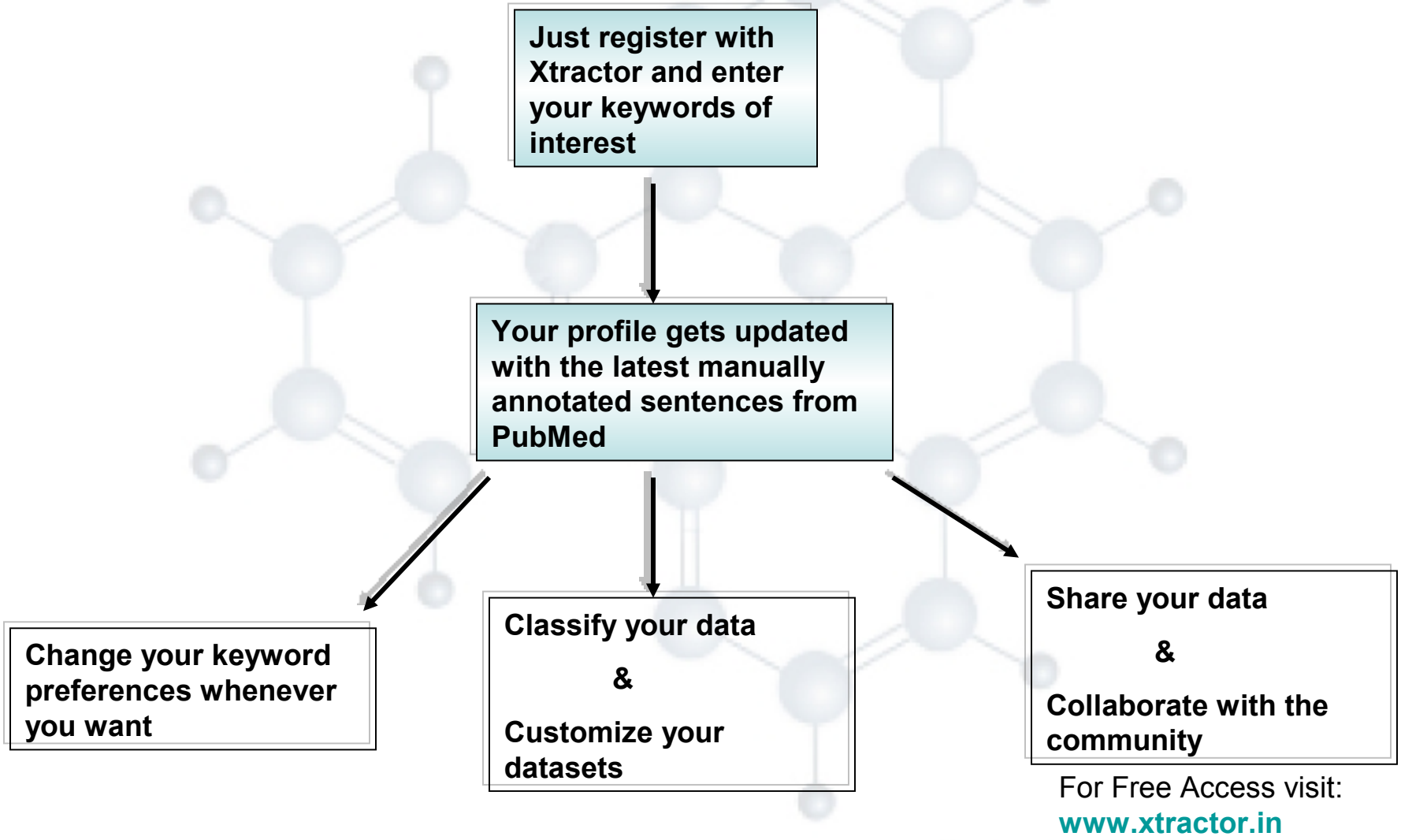
NLP Vs Manual Curation:



We performed an analysis comparing "standard NLP" vs manual curation efforts. It was found that there are 12-35% false positive picks with the NLP in comparison to the manual approach that we employed.

For Free Access visit:
www.xtractor.in

The Xtractor way



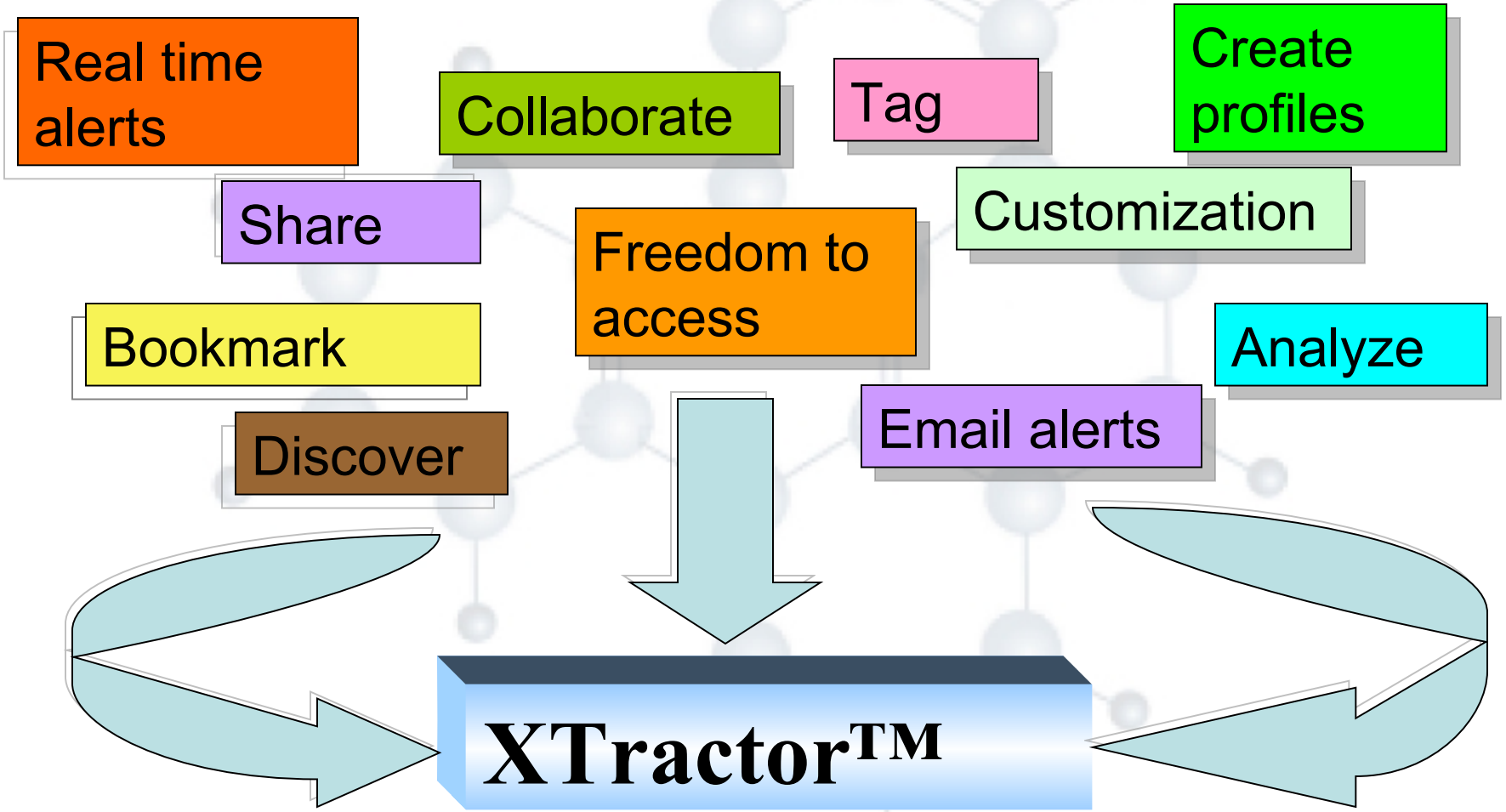
For Free Access visit:
www.xtractor.in

Product highlights

- **Manually validated** and **expert annotated** scientific facts
- **Sharing** of stored facts and articles
- **Discovery and creation** of newer relations through XTractor Community
- **Bookmarks of** favorite abstracts & annotated sentences
- **Classify and create** your own datasets of annotated facts
- **Access** your datasets from anytime, anywhere
- **Email** alerts

For Free Access visit:
www.xtractor.in

XTractor features



For Free Access visit:
www.xtractor.in

XTractor Advantage

MOLECULAR CONNECTIONS
Knowledge Solutions for Accelerated Research

Store your choice of sentences and create your own customized datasets for your experiments

Analyze relations between abstracts and discover new relations from the voluminous data pool

Highly accurate and more comprehensive than Natural Language Processing engines

For Free Access visit:
www.xtractor.in

XTractor Advantage

Stay current with the hot areas of research and publications

Xtractor helps you to find complex biological relations

Sentences are categorized based on the relationships they exhibit: e.g. protein- drug or gene- knockouts, etc.

For Free Access visit:
www.xtractor.in

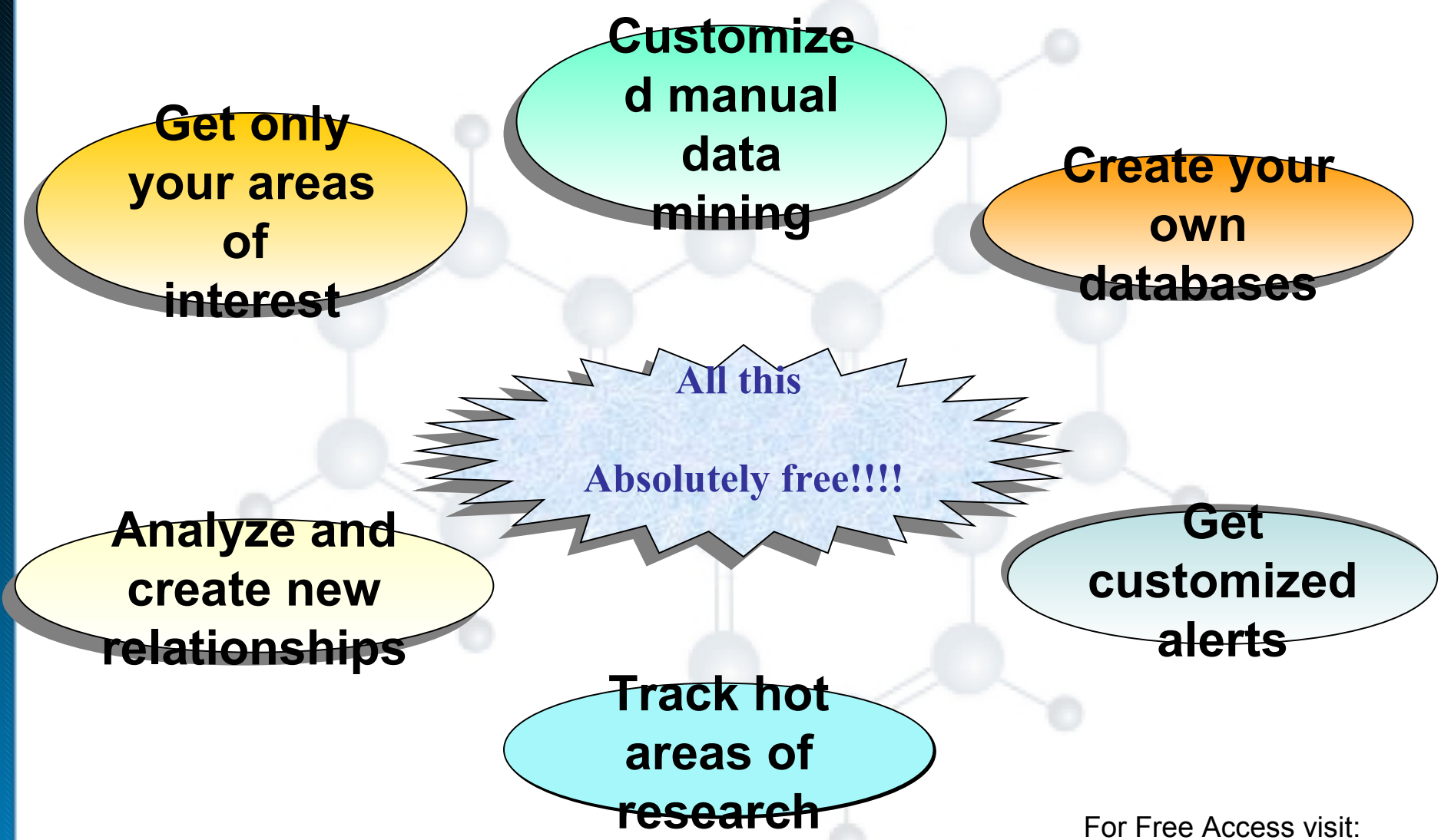
XTractor Advantage

Change your keywords, as your areas of interest change - at no costs

Highly cost effective – pay only for your choice of keywords, presently Absolutely free !!!

For Free Access visit:
www.xtractor.in

Xtractor snapshot



For Free Access visit:
www.xtractor.in

What's next?

Upcoming features…

For Free Access visit:
www.xtractor.in

Basic level search

User searches based on ... the following

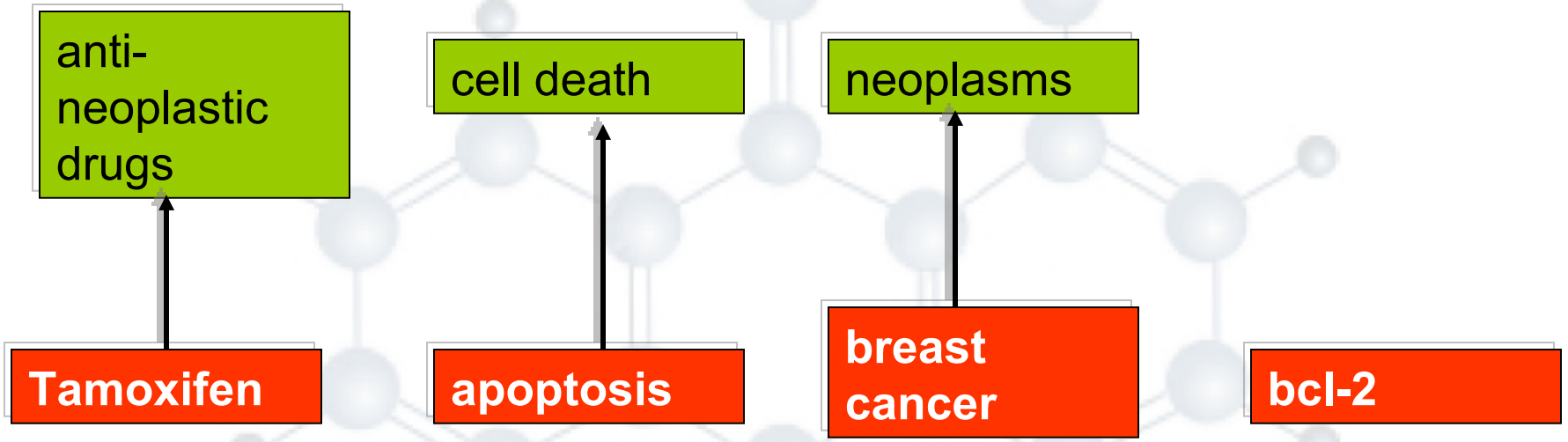
Tamoxifen **apoptosis** **breast cancer** **bcl-2**

Tamoxifen-induced apoptosis in breast cancer cells relates to down-regulation of bcl-2, but not bax and bcl-X(L), without alteration of p53 protein levels.

For Free Access visit:
www.xtractor.in

Search 1 level up the ontology

Moving levels up the ontology improves retrieval



Tamoxifen-induced apoptosis in breast cancer cells relates to down-regulation of bcl-2, but not bax and bcl-X(L), without alteration of p53 protein levels.

Additional sentences are picked:

Our data show that Bcl-2 can protect antiestrogen-resistant breast cancer cells from cisplatin-induced cell death, indicating that the reduced expression of Bcl-2 in the antiestrogen-resistant cells plays a role in sensitizing the cells to cisplatin treatment.

For Free Access visit:
www.xtractor.in

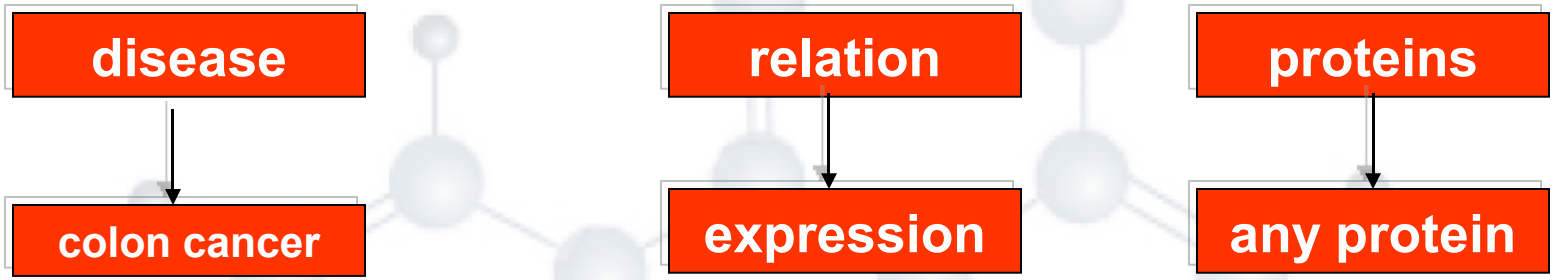
Broad searches

any drug | apoptosis | kinase

Show me “any drug” which is involved in ‘apoptosis” and inhibits a “kinase”

Retrieves large number of hits which could be used for next level screening

Using relationships to traverse ontology



- These findings support evidence of the enhanced **RELMbeta expression** in **colon cancer** patients and suggest that further investigation is warranted to explore the role of RELMbeta in colon cancer.
- **Proapoptotic Bad and Bid Protein Expression Predict Survival in Stages II and III Colon Cancers.**
- In conclusion, we show how serotonin regulates angiogenesis in s.c. **colon cancer** allografts by influencing **MMP-12 expression** in tumor-infiltrating macrophages, thereby affecting the production of circulating angiostatin.
- Somatostatin decreases **COX-2 expression** and function in **colon cancer** cells via activation of sst(3) or sst(5) receptors, and these effects contribute to the inhibitory action of somatostatin on cell proliferation.

Bibliographic search module

- Search by author
- Search by institutions
- Search by publication dates
- **Also since we have manually annotated data each author/institution could be searched by proteins, process, disease and drugs.**

For Free Access visit:
www.xtractor.in

Search though archived data

- This is for users who cannot maintain their profiles over the long run

OR

- Clients with immediate data requirements

- Perform a search on archived data (filter with a date range)

For Free Access visit:
www.xtractor.in

XTractor as a web service

- Xtractor would also provide the user with the ability to export XML data as a web service
- This would enable users to create their own datasets of the XML dumps at their end
- Also the data could be integrated into proprietary systems

For Free Access visit:
www.xtractor.in

Order a Report

- The XTractor team would classify the archived data into customized reports
- Reports on breast cancer- (disease conditions- target genes, therapies etc), OR say reports on toxicity of a drug etc..
- We may have these reports showcased on Xtractor and sold for a flat price.

For Free Access visit:
www.xtractor.in

Enhanced experiences sentences Literature Discover Create desktop extracted facts Provision change keywords Daily updates changing research preferences updates stored profile **manually annotated** published Classify create datasets annotated facts keywords share reading abstracts **XTractor™** analyzing literature **Access datasets** user preference anytime, anywhere newer relations scientific **Absolutely free** facts classified annotated Tag sentences Community favorite abstracts **Highly accurate** **collaborate** fellow scientists faster service

For FREE registration :www.xtractor.in

For Free Access visit:
www.xtractor.in