

The case...

- **Manos** is a PhD student in BSRC A. Fleming, GR, starting his research on microRNAs.



The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Prof. says: "they bind to complementary sequences on target messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Target prediction???
Need to find some papers!



Prof. says: "they bind to complementary sequences on **target** messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Target prediction???
Need to find some papers!



Prof. says: "they bind to complementary sequences on **target** messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".



Found some papers in Google!

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Target prediction???
Need to find some papers!



Prof. says: "they bind to complementary sequences on **target** messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".



Found some papers in Google!



DIANA microT
TargetScan

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Target prediction???
Need to find some papers!



Prof. says: "they bind to complementary sequences on **target** messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".



Found some papers in Google!



Good Idea!!
DIANA microT
TargetScan

The case...



MicroRNAs (miRNAs) are short ribonucleic acid (RNA) molecules, on average only 22 nucleotides.



Computational verified
Target prediction???
Need to find some papers!
Experimentally verified



Prof. says: "they bind to complementary sequences on **target** messenger RNA transcripts (mRNAs), usually resulting in translational repression and gene silencing".

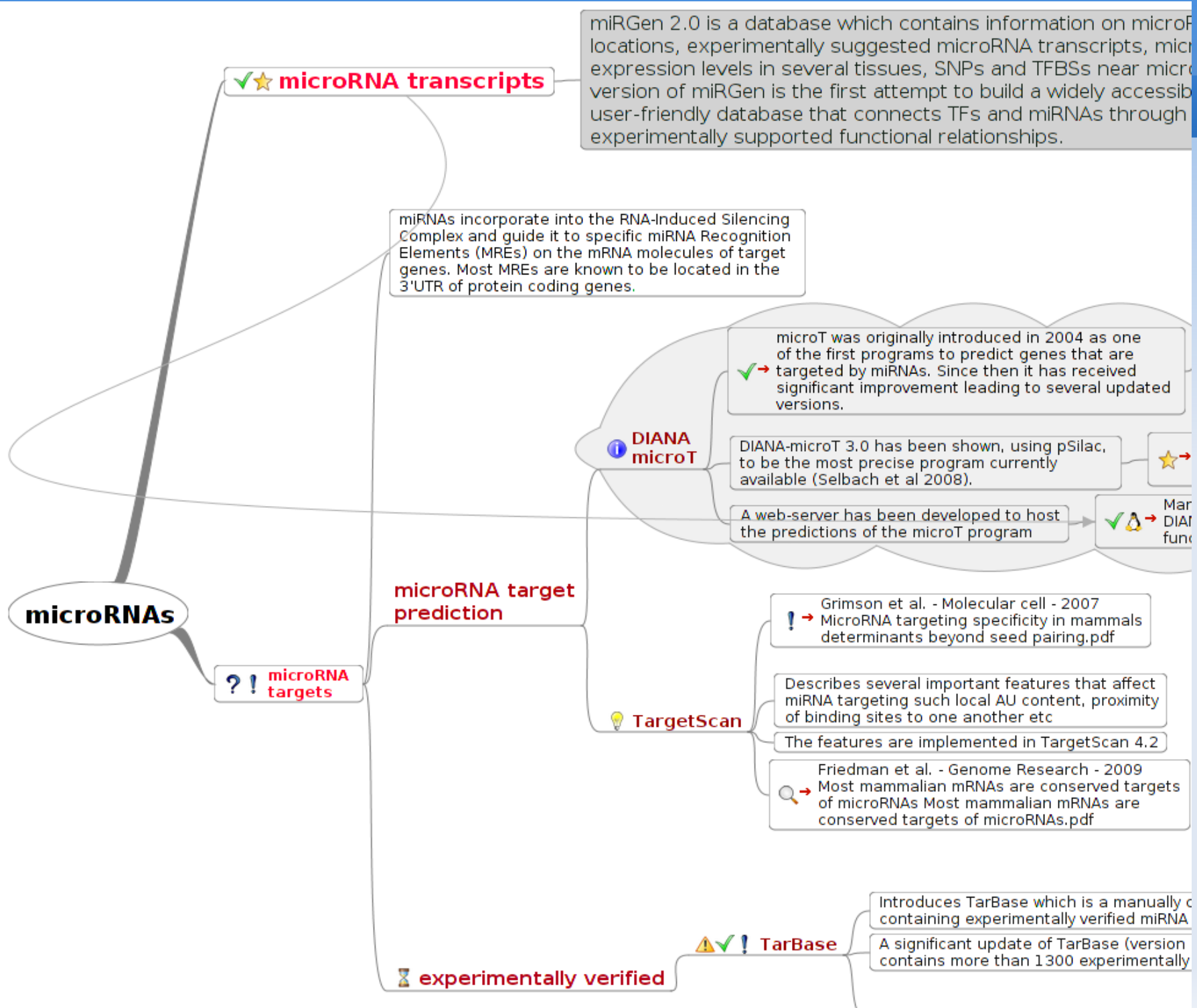


Found some papers in Google!



Good Idea!!
DIANA microT
TargetScan

Find more papers, write notes, generate ideas, re-structure the domain...



Mindmaps

- Graphical representations of **elements**
 - concepts, ideas, notes, tasks, or other items related to a topic of study.
- **Mindmapping elements** are organized in hierarchical branches.
- However, everything is built around a **central topic**.
 - Elements are arranged in a non-linear fashion.
- This encourages a **brainstorming approach** to planning and organizational tasks, and idea generation.

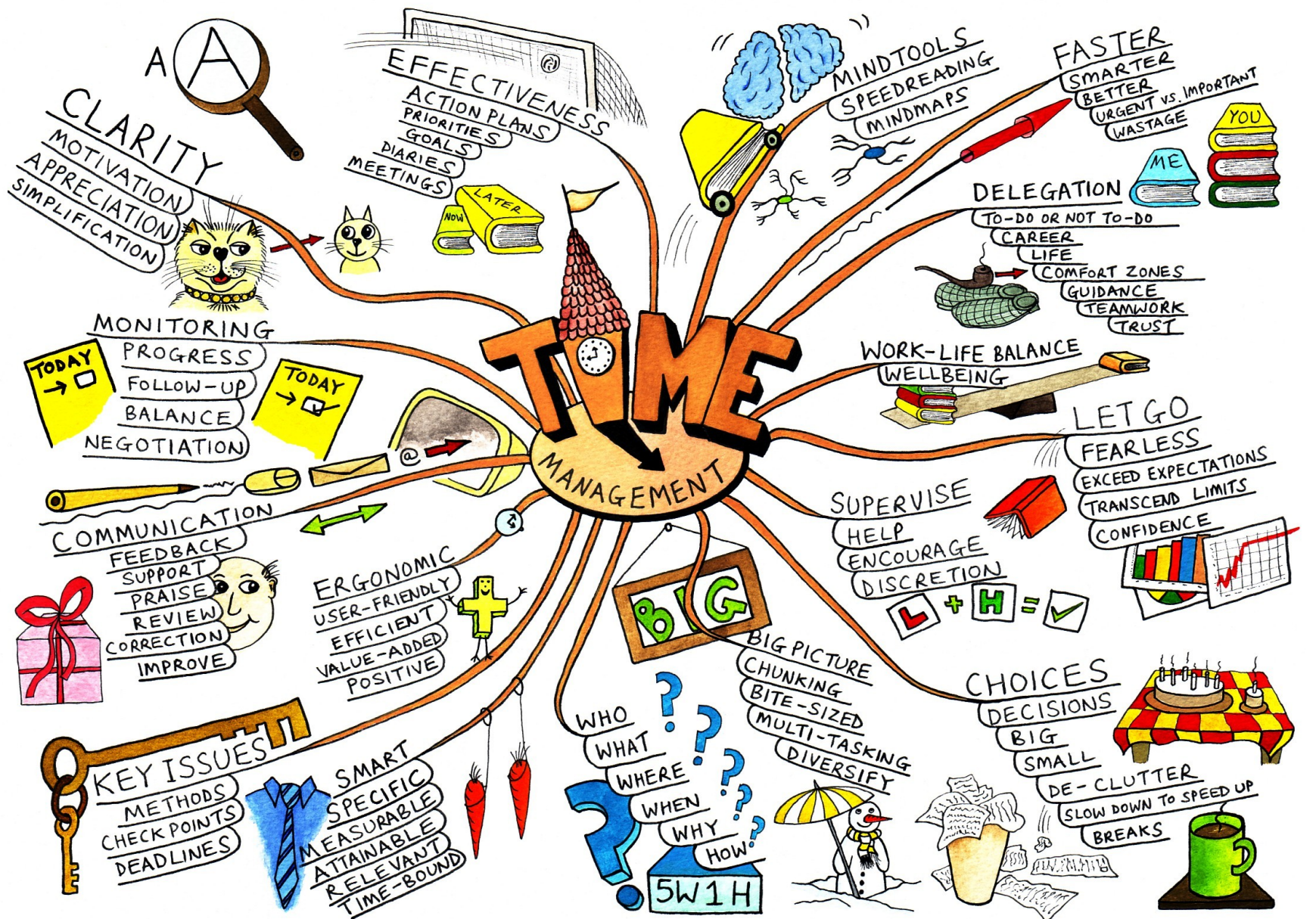
Mindmaps

- Mindmap usage
 - project and task management
 - links to files, executables, source of information,
 - knowledge base
 - organization (notes, refs), essay writing and brainstorming.
- Mindmaps vs Semantic Networks, TopicMaps
 - No explicit structured elements to model relationships. **Radial hierarchies.**
 - **Visual mnemonics:** colors, icons and informal visual representations.

Mindmaps

- Visual methods like **mindmaps** have been used for centuries
 - learning
 - problem solving
 - recording knowledge
 - visual thinking

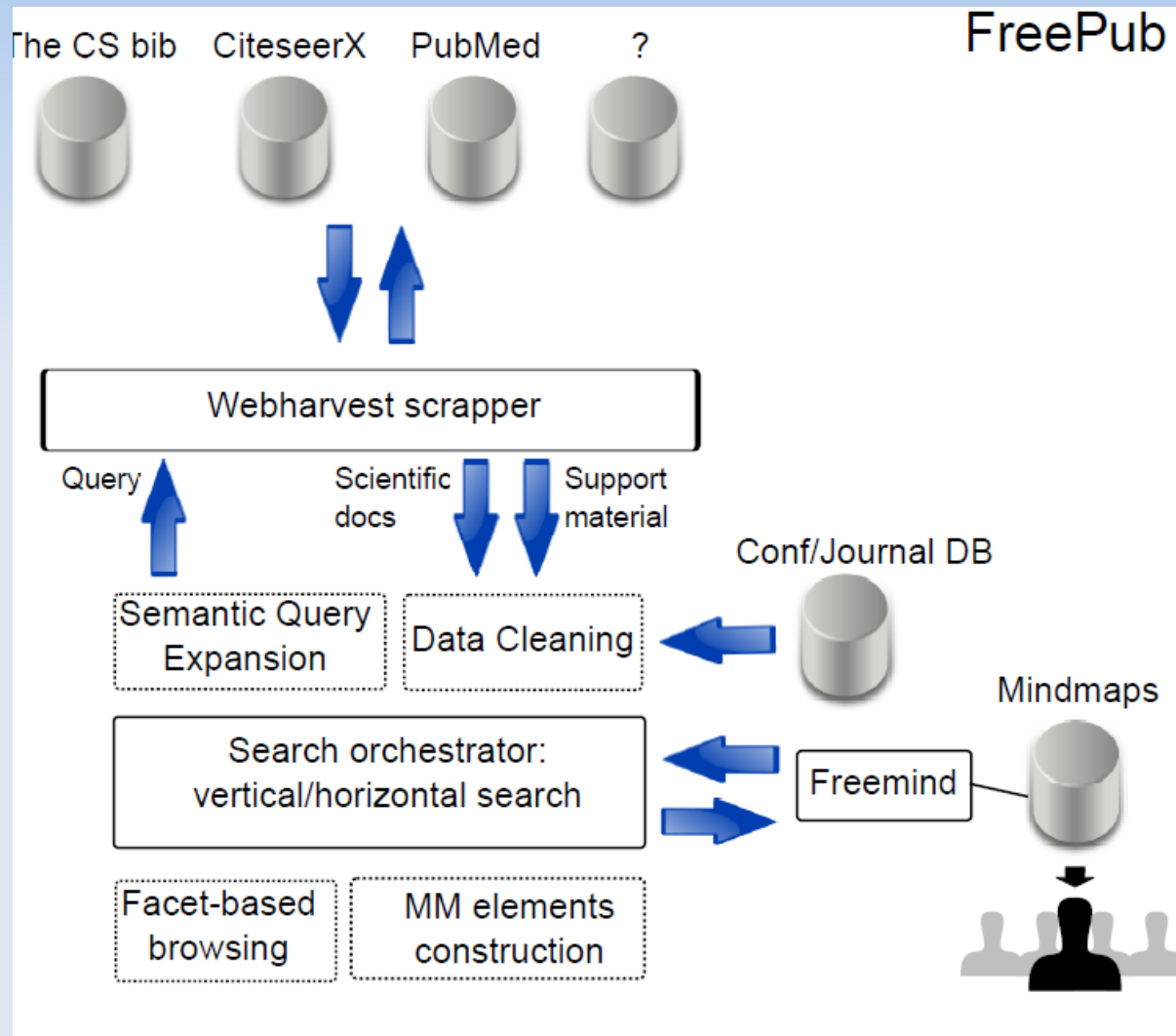
Mindmap in paper!



FreePub

- **FreePub, a creativity support tool** to collect and organize scientific material using **mindmaps**.
- **Creativity cycles:**
 - set up domain of interest using mindmaps,
 - browse mindmaps and launch search tasks to gather relevant documents from several data sources,
 - identify supporting material for those documents (e.g., blog posts, presentations),
 - import and organise all retrieved information in mindmaps.

FreePub architecture



Searching facilities

- **Vertical search**

- Retrieve relevant documents from several data sources.
- Vertical = tailored to certain types of information resources (e.g., DBLP, CiteseerX, PubMed, ...).

- **Horizontal search**

- Retrieve supporting material to retrieved docs.
- Horizontal = generic search engines (e.g., Google).
- Examples: doc itself, abstract, slide presentation, blog entries.

Vert. search facilities: problems

- **Wrapping?**
- Data sources should be easily integrated.
- Web scrappers
 - Capture data source search capabilities.
 - Simplify Web information extraction.
- **WebHarvest** (<http://web-harvest.sourceforge.net/>)
 - Provides several types of processors (e.g., html-to-xml, xpath, etc).
 - Users define a **sequence of extraction operations** on Web pages based on **path expressions**.

Vert. search facilities: problems

- **Cleaning?**
- Ambiguities in conf/journal names.
- Maintain a list of (name, acronym) pairs.
- Use **string edit distance** to determine similarities.
 - Min number of edit operations (INS, DEL, REPL) needed to transform one string into the other.
 - $L(\text{'VLDD'}, \text{'VLDB Conf'}) = 6$: replace `D' with `B', and insert ` ', `C', `o', `n', `f'.
 - 'Very Large Database Conf' and 'VLDB Conf, both are matched to ('Very Large Database Conference', 'VLDB) catalog entry.

Vert. search facilities: problems

- **Duplicate elimination?**
- Follow an entity resolution blocking technique.
 - Divide objects into groups and only compare objects within the same group, to avoid redundant comparisons.
 - Records in different groups are unlikely to match.
- Partition the result list of each data source into groups (key = date for each group)
 - Maintain a (key, {list of docs}) structure
- Check pairs (o1, o2) only for o's that share the same key (date).

Hor. search facilities: problems

- **Support material?**
- Doc itself, e.g.:
 - title AND filetype:pdf OR doc
 - Parse the doc to identify title
- Abstract, e.g.:
 - Parse doc for "Abstract" tag.
 - Many data sources do provide abstract!
- Slide presentations, e.g.:
 - Title AND filetype:ppt
 - Parse doc for certain terms (e.g., outline) (under impl).

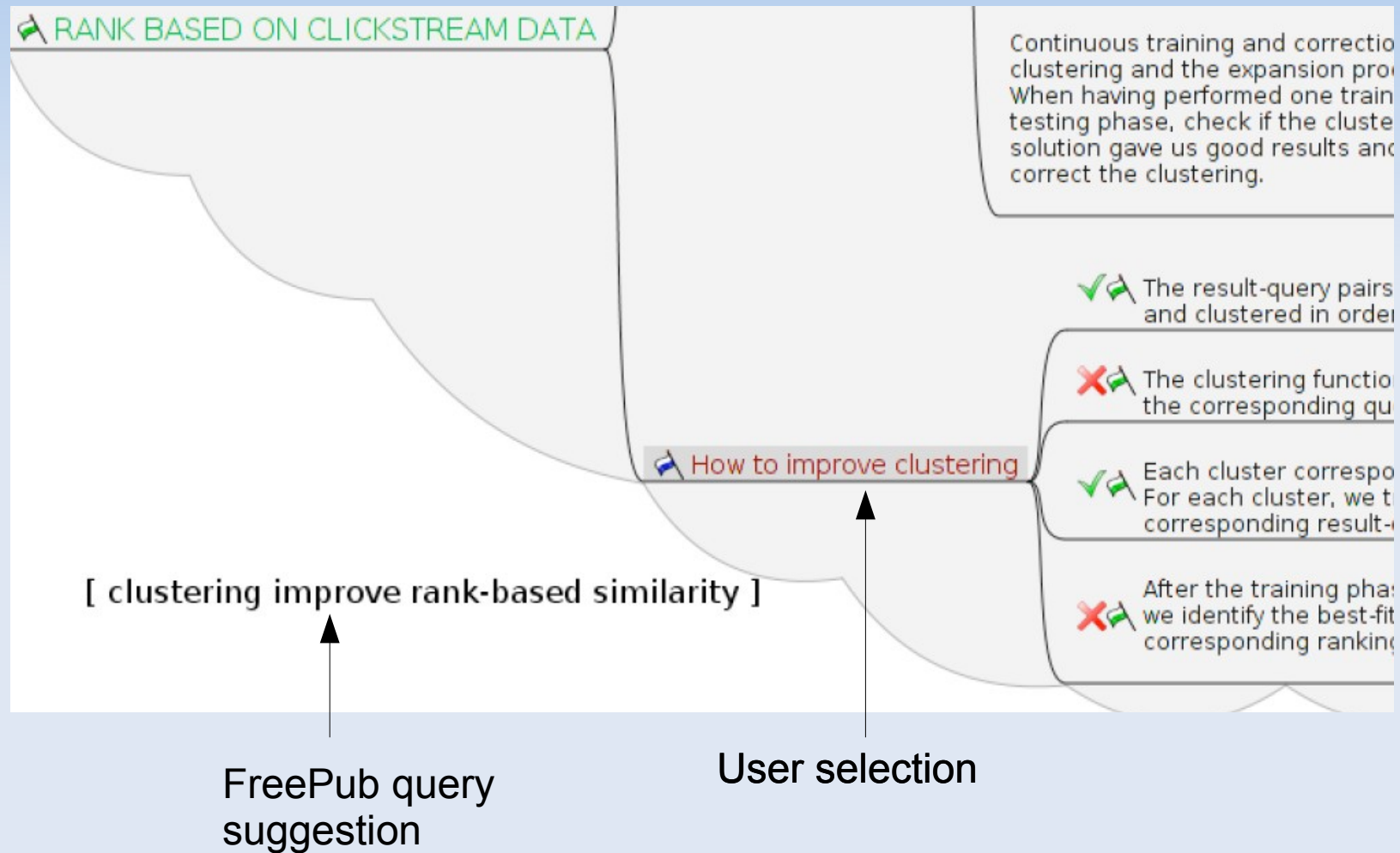
Hor. search facilities: problems

- **Support material?**
- Title AND author names
- Google Blog Search
- Technorati (<http://technorati.com/>) (under impl.)

Using FreePub

- User selects mindmap elements.
- Freepub **extracts keywords** from these elements.
- Metasearch orchestration.
 - Vertical search: find relevant documents.
 - Horizontal search: find supporting material.
- Two modes for keyword extraction
 - **Normal** mode.
 - **Semantic query expansion** mode.

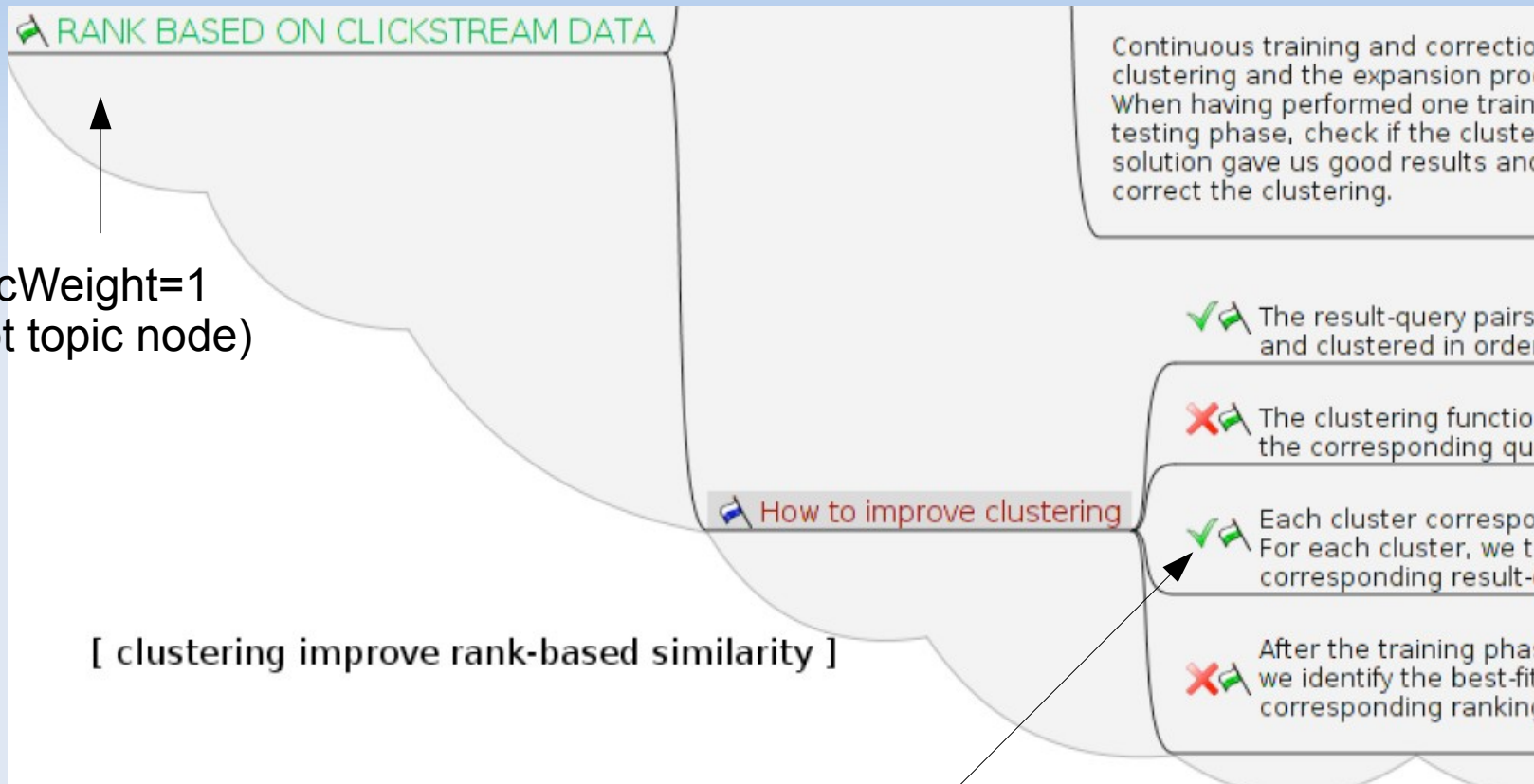
Semantic query expansion



Semantic query expansion

- All elements in the **neighbourhood** of user-selected elements are considered as documents.
 - The level of neighbourhood is user defined.
 - E.g., level 1 means that the neighbourhood of a selected element includes only **directly adjacent** nodes.
- To each document **d**, we assign weights **docWeight_d** according to the **type** of corresponding elements.

Semantic query expansion



DocWeight=0.8
(subtopic node)

Semantic query expansion

- How to determine expansion terms?
 - Doc terms are cleaned (i.e., punctuation and stopwords are removed).
 - **docSize(d)** = number of terms for document **d**.
 - Compute **freq(t,d)**: occurrences of **t** in each **d** (i.e., term frequency), and **docFreq(t)**: num of docs containing **t**.
 - Compute **w(d,t)** = $\text{freq}(t,d) * \text{docFreq}(t) * \text{docWeight}(d) / \text{docSize}(d)$.
 - Term weight **W(t)**: the average of **w(d,t)**'s.

Semantic query expansion

- Terms are sorted according to $W(t)$.
- Use the TOP-K terms to expand the initial query.

Some screenshots

Nature Precedings : doi:10.1038/npre.2010.5432.1 : Posted 23 Dec 2010

The image displays two overlapping screenshots. On the left is a mind map titled 'target' with two main branches: 'DIANA-microT' and 'TargetScan'. 'DIANA-microT' includes sub-branches like 'DIANA to be available' and 'A web the p'. 'TargetScan' includes sub-branches like 'Describe miRNA target of binding', 'The feat', and 'Frie Mos of m con'. A note at the bottom of the mind map says 'Naive Bayes methods'. On the right is a screenshot of a search engine interface titled 'Freepub v0.1 beta'. It features a 'Search Engines' section with radio buttons for 'CCSB', 'CiteSeerX', and 'PubMed'. The 'Search Query Options' section includes a search query box containing 'methods naive bayes target microrna prediction', checkboxes for 'Include Node Text' and 'Activate Semantic Query Expansion', and an 'Extra Keywords' field. Below these is a table of search results.

Title	Date	Forum	Authors
Na\ive Bayes for microRNA target prediction...	2007	Bioinformatics	Malik Yousef, Segun Ju...
Learning from positive examples when the n...	2008		Yousef Malik, Jung Seg...
Structural conservation versus functional div...	2008		Martin Kircher, Christo...
Structural conservation versus functional div...	2008		Kircher Martin, Bock Ch...
Combining multi-species genomic data for mi...	2006	Bioinformatics	Malik Yousef, Michael N...
A comparison of event models for Naive Bay...	1998		Andrew McCallum, Kam...
Empirical Analysis of Predictive Algorithms fo...	1998		John S. Breese, David ...
Wrappers for Feature Subset Selection	1997	Artificial Intelligence	Ron Kohavi, George H. ...
BLEU: a Method for Automatic Evaluation of M...	2002		Kishore Papineni, Sali...
Text Classification from Labeled and Unlabel...	1999	Machine Learning	Kamal Nigam, Andrew ...
A Comparative Study on Feature Selection in ...	1997		Yiming Yang, Jan O. Pe...
Online Learning with Kernels	2003		Jyrki Kivinen, Alexander...
Text Categorization with Support Vector Mach...	1997		Thorsten Joachims, Fac...
Instance-based learning algorithms	1991	Machine Learning	David W. Aha, Dennis K...
A tutorial on particle filters for online nonline...	2002	IEEE Transactions on Sign...	M. Sanjeev Arulampala...

Some screenshots

The screenshot shows the Freepub v0.1 beta application window. The interface is divided into several sections:

- Options:** Includes a search bar with a dropdown set to '5', a 'Search' button, a 'Cancel' button, and an 'Import in mindmap' button.
- Support Material:** A section on the right containing a 'Load Selected' button and a list of search results.
- Table:** A table with columns for 'Authors', 'Extra', and 'Import'. The second row is highlighted in orange.
- Document View:** A detailed view of a document, including an abstract, document title, presentations, and a blog entry.

Authors	Extra	Import
Malik Yousef, Segun Ju...	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Yousef Malik, Jung Seg...	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Martin Kircher, Christo...	<input type="checkbox"/>	<input type="checkbox"/>
Kircher Martin, Bock Ch...	<input type="checkbox"/>	<input type="checkbox"/>
Malik Yousef, Michael N...	<input type="checkbox"/>	<input type="checkbox"/>
Andrew McCallum, Kam...	<input type="checkbox"/>	<input type="checkbox"/>
John S. Breese, David ...	<input type="checkbox"/>	<input type="checkbox"/>
Ron Kohavi, George H. ...	<input type="checkbox"/>	<input type="checkbox"/>
Kishore Papineni, Sali...	<input type="checkbox"/>	<input type="checkbox"/>
Kamal Nigam, Andrew ...	<input type="checkbox"/>	<input type="checkbox"/>
Yiming Yang, Jan O. Pe...	<input type="checkbox"/>	<input type="checkbox"/>
Jyrki Kivinen, Alexander...	<input type="checkbox"/>	<input type="checkbox"/>
Thorsten Joachims, Fac...	<input type="checkbox"/>	<input type="checkbox"/>
David W. Aha, Dennis K...	<input type="checkbox"/>	<input type="checkbox"/>
M. Sanjeev Arulampala...	<input type="checkbox"/>	<input type="checkbox"/>

Abstract:
Abstract{\par} {\par}Background{\par} {\par}The application of machine learning to classification problems that depend only on positive examples is gaining attention in the computational biology community. We and others have described the use of two-class machine learning to identify novel miRNAs. These methods require the generation of an artificial negative class. However, designation of the negative class can be problematic and if it is not properly done can affect the performance of the classifier dramatically and/or yield a biased estimate of performance. We present a study using one class

Document: [More](#)

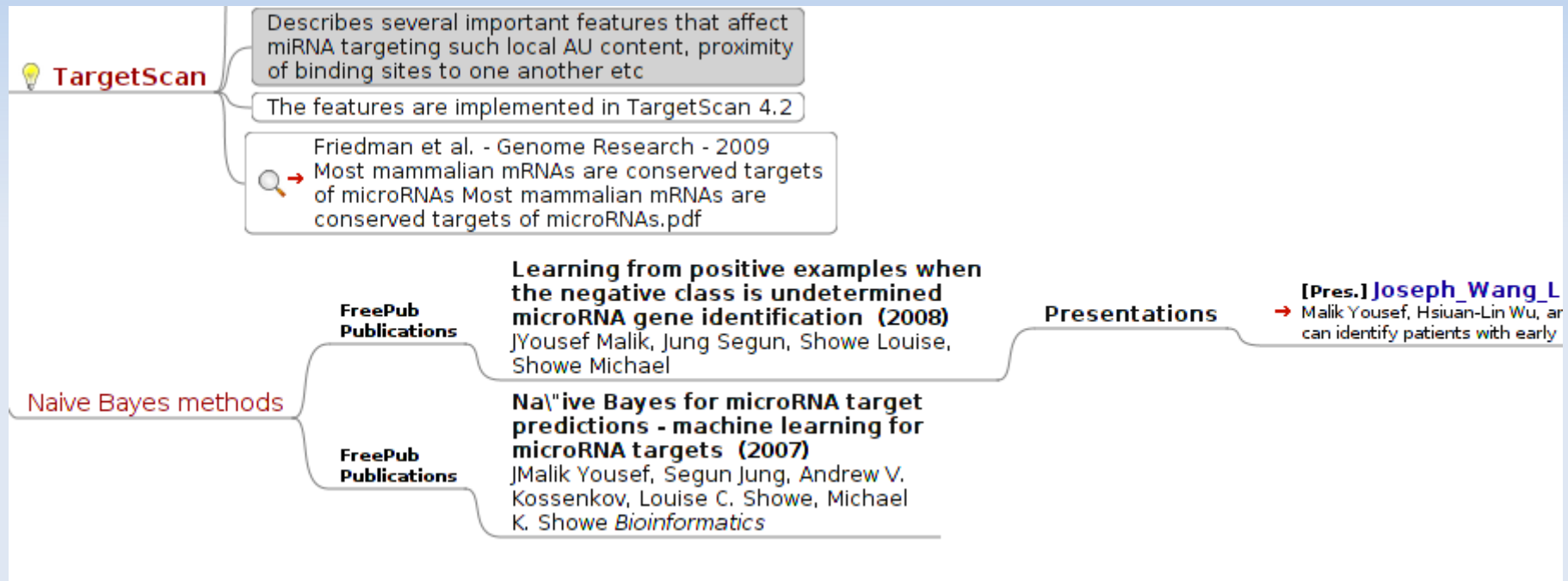
Presentations: [More](#)

- PowerPoint Presentation**
Malik Yousef, Hsiuan-Lin Wu, and Michael Nebozhyn ...Support tab-delimited data file with multiple gene annotations as well as multiple sample annotations. Analytical Algorithms ...VISDA UML ClassDiagram. edu.vt.cbil.visda.view.
http://www.cbil.ece.vt.edu/software/cabiq_visda_demo.ppt
- Joseph Wang Louise Showe VISDA ppt - PowerPoint Presentation**
Malik Yousef, Hsiuan-Lin Wu, and Michael NebozhynIdentify GeneExpression patterns in peripheral blood samples that can identify patients with early lung cancer ... Cases within classes are expected to be heterogeneous
https://cabiq.nci.nih.gov/archive/ICR/Meetings/index_html/2006_Janu

Blogs: [More](#)

- Algorithms for Molecular Biology | Full text | Learning from ...**
In a recent publication we described a two-class machine learning approach for miRNA prediction using the naïve Bayes classifier [13]. Four criteria were used to select a pool of negative examples from candidate stem loops: stem length out of the range
<http://www.almob.org/content/3/1/2>

Some screenshots



FreePubCollecting and Organizing Scientific Material Using Mindmaps

Theodore Dalamagas, IMIS Institute /'Athena' R.C., Athens, Greece

Tryfon Farmakakis, School of Informatics, University of Edinburgh, Scotland

Manolis Maragkakis, BSRC 'A. Fleming', Greece

Artemis G. Hatzigeorgiou, BSRC 'A. Fleming', Greece

<http://web.imis.athena-innovation.gr/projects/mm/>

