

# Ranked retrieval of annotated models

*Combine 2010*



*Ron Henkel*



*Dagmar Waltemath*

## Models

- What is a model?
- What is a model for me?
  - a piece of XML code
  - (semi) structured
  - annotated
  - MIRIAM compliant



```
<species id="C_p" sboTerm="SBO:0000247">
  <annotation>
    <rdf:Description rdf:about="C_p">
      <bqbiol:is>
        <rdf:Bag>
          <rdf:li rdf:resource=
            "urn:miriam:obo.chebi:CHEBI%3A27732"/>
        </rdf:Bag>
      </bqbiol:is>
      <bqbiol:is>
        <rdf:Bag>
          <rdf:li rdf:resource=
            "urn:miriam:kegg.compound:C07481"/>
        </rdf:Bag>
      </bqbiol:is>
    </rdf:Description>
  </annotation>
</species>
```

## What do we know about a model?

- Encoded in the model itself:
  - Who submitted a model, who encoded it?
  - When was the model created or modified?
  - Name, id and notes of constituents
- Available via external resources
  - What publication is related to a model?
  - Biological/chemical description of constituents

## Assumptions

- Information encoded in a model itself is not sufficient for searching `sp1, re_2`
- Users want to search using meaningful biological terms `caffeine OR 3,7-Dihydro-1,3,7-trimethyl-1H-purin-2,6-dion`
- Users don't want to type in complex queries `le Novère`
- Aspects of a model might have a different importance

## Dimensions and weights

| # | dimension      | importance | description                                     |
|---|----------------|------------|-------------------------------------------------|
| 1 | Administrative | none       | ids, file name, version, formalism...           |
| 2 | Person         | medium     | creator, encoder, submitter, publication author |
| 3 | Dates          | low        | creation and modification date                  |
| 4 | Publication    | high       | title, abstract, full-text, journal             |
| 5 | Constituents   | very high  | compartment, species, reaction                  |
| 6 | User generated | very high  | keywords, tags, remarks, changes                |

- The concept is abstract and can be applied to different model formalisms
- Depending on the formalism the dimensions are refined into features
- Our example implementation uses SBML

## Model & Semantic index

- Model index
  - contains information encoded in the model itself
  - separates information encoded in a model into logic parts
- Semantic index
  - contains information represented by URIs
  - maps the relation between qualifier, URI and textual representation

## Create the index

```
<species id="C_p" sboTerm="SBO:0000247">
  <annotation>
    <rdf:Description rdf:about="C_p">
      <bqbiol:is>
        <rdf:Bag>
          <rdf:li rdf:resource=
            "urn:miriam:obo.chebi:CHEBI%3A27732"/>
        </rdf:Bag>
      </bqbiol:is>
      <bqbiol:is>
        <rdf:Bag>
          <rdf:li rdf:resource=
            "urn:miriam:kegg.compound:C07481"/>
        </rdf:Bag>
      </bqbiol:is>
    </rdf:Description>
  </annotation>
</species>
```

### Model Index

| BM-ID             | species | speciesURI                 | ... |
|-------------------|---------|----------------------------|-----|
| BIOMD<br>00000241 | C_p     | Chebi:27732<br>Kegg:C07481 | ... |

### Semantic Index

| URI                        | qualifier | content                                                                             |
|----------------------------|-----------|-------------------------------------------------------------------------------------|
| Chebi:27732<br>Kegg:C07481 |           | caffeine, ...<br>C <sub>8</sub> H <sub>10</sub> N <sub>4</sub> O <sub>2</sub> , ... |

model

species

URI

qualifier

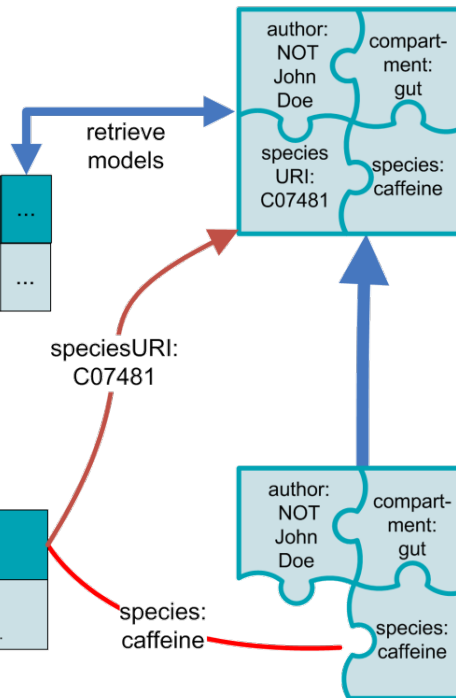
# Query the index

## Model Index

| BM-ID         | species | speciesURI                 | ... |
|---------------|---------|----------------------------|-----|
| BIOMD00000241 | C_p     | Chebi:27732<br>Kegg:C07481 | ... |

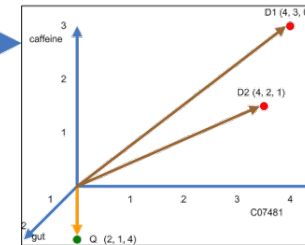
## Semantic Index

| URI                        | qualifier | content                                                                             |
|----------------------------|-----------|-------------------------------------------------------------------------------------|
| Chebi:27732<br>Kegg:C07481 |           | caffeine, ...<br>C <sub>8</sub> H <sub>10</sub> N <sub>4</sub> O <sub>2</sub> , ... |



relevant models

## Vector Space



ranked list of relevant models

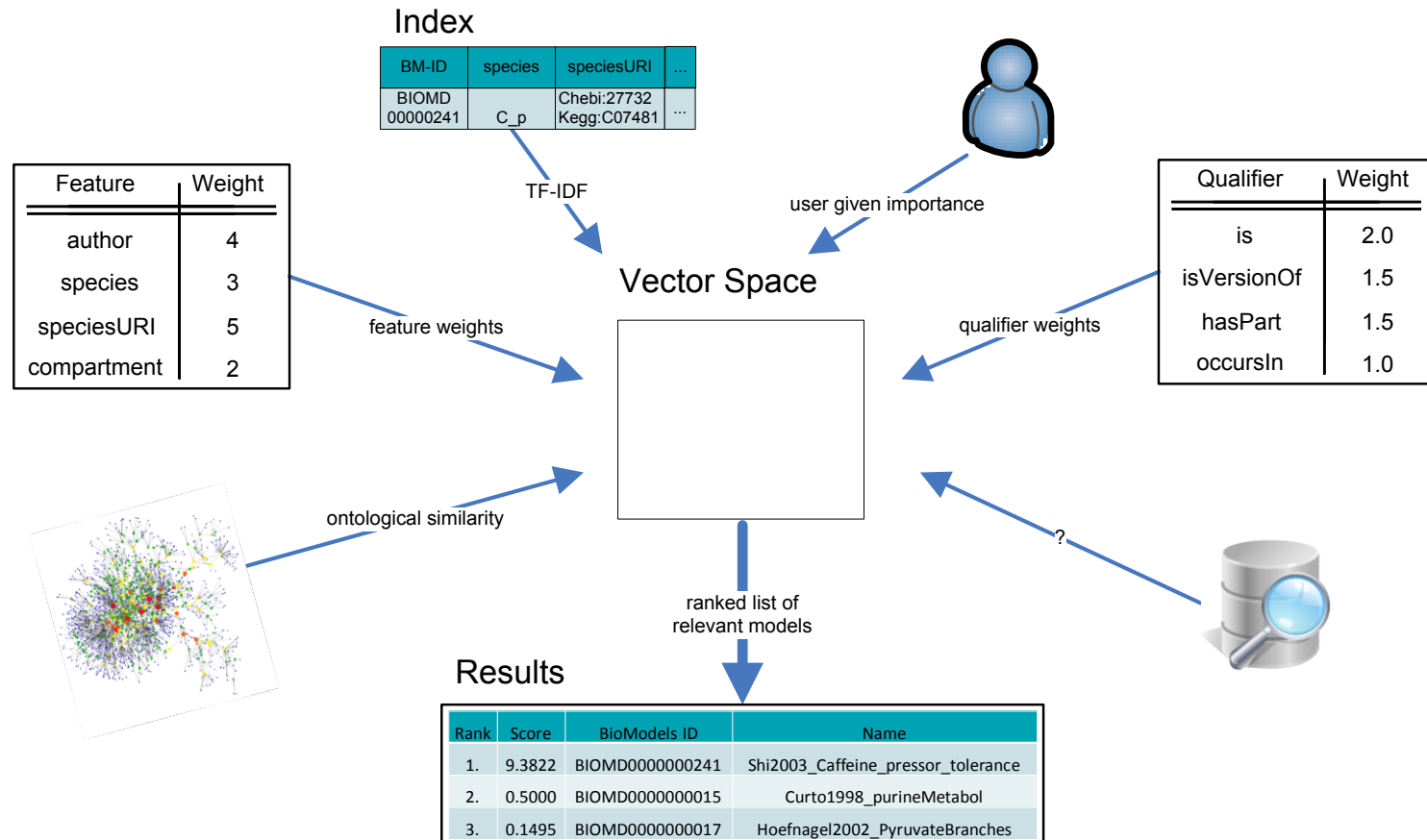
## Results

| Rank | Score  | BioModels ID    | Name                               |
|------|--------|-----------------|------------------------------------|
| 1.   | 9.3822 | BIOMD0000000241 | Shi2003_Caffeine_pressor_tolerance |
| 2.   | 0.5000 | BIOMD0000000015 | Curto1998_purineMetabol            |
| 3.   | 0.1495 | BIOMD0000000017 | Hoefnagel2002_PyruvateBranches     |

Models by non-bogus authors describing the effect of caffeine in human's digestive tract when drinking a cup of coffee?



# Rank the results



## Search for models by **keywords**

|                |                                                                |                                  |                                         |           |                                                                                                                                                 |
|----------------|----------------------------------------------------------------|----------------------------------|-----------------------------------------|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Person:        | <input type="text" value="John Doe"/>                          | <input type="button" value="+"/> | <input type="button" value="must not"/> | occur in: | <input type="button" value="submitter"/><br><input type="button" value="encoder"/><br><input checked="" type="button" value="author"/>          |
| SBML Elements: | <input type="text" value="gut"/>                               | <input type="button" value="+"/> | <input type="button" value="should"/>   | occur in: | <input type="button" value="species"/><br><input checked="" type="button" value="compartment"/><br><input type="button" value="reaction"/>      |
| Resource:      | <input type="text" value="caffeine"/>                          | <input type="button" value="+"/> | <input type="button" value="must"/>     | occur in: | <input type="button" value="ALL"/><br><input type="button" value="modelURI"/><br><input checked="" type="button" value="speciesURI"/>           |
|                | limit to qualifier: <input type="text" value="bqbiol:is"/>     |                                  |                                         |           |                                                                                                                                                 |
| Dates:         | <input type="text" value="01/01/2009"/> - <input type="text"/> | <input type="button" value="+"/> | <input type="button" value="should"/>   | occur in: | <input checked="" type="button" value="ALL"/><br><input type="button" value="modificationDate"/><br><input type="button" value="creationDate"/> |

## Search for models by **keywords**

- `-[PERSON]: John Doe ONLY IN author`
- `[SBML-ELEMENT]: gut ONLY IN compartment`
- `+ [RESOURCE]: caffeine TRANSFORMED TO (urn:miriam:obo.chebi:chebi:27732^0.8187308  
urn:miriam:kegg.compound:c07481^0.67032003  
urn:miriam:kegg.compound:c00385^0.5488116  
ONLY IN speciesURI / LIMITED TO QUALIFIER bqbiol:is`
- `[DATE]: FROM 01/01/2009`

3 Curated Models returned.

| Rank                          | BioModels ID                    | Name                               | Publication ID           | Last Modified             |
|-------------------------------|---------------------------------|------------------------------------|--------------------------|---------------------------|
| 1. ( <a href="#">9.3822</a> ) | <a href="#">BIOMD0000000241</a> | Shi2003_Caffeine_pressor_tolerance | <a href="#">8422743</a>  | 2010-01-11T16:04:55+00:00 |
| 2. ( <a href="#">0.5000</a> ) | <a href="#">BIOMD0000000015</a> | Curto1998_purineMetabol            | <a href="#">9664759</a>  | 2009-07-03T08:06:44+00:00 |
| 3. ( <a href="#">0.1495</a> ) | <a href="#">BIOMD0000000017</a> | Hoefnagel2002_PyruvateBranches     | <a href="#">11932446</a> | 2010-01-18T10:48:56+00:00 |


## Search for models by **example model**

The search returned 453 models.

Processed query was:

- Models similar to BIOMD0000000171

☐ 241 Curated Models returned.

| Rank  | BioModels ID                    | Name                    | Publication ID           | Last Modified             |
|----------------------------------------------------------------------------------------|---------------------------------|-------------------------|--------------------------|---------------------------|
| 1. ( <a href="#">2.1102</a> )                                                          | <a href="#">BIOMD0000000171</a> | Leloup1998_CircClock_LD | <a href="#">9486845</a>  | 2010-02-08T10:33:14+00:00 |
| 2. ( <a href="#">1.3009</a> )                                                          | <a href="#">BIOMD0000000021</a> | Leloup1999_CircClock    | <a href="#">10366496</a> | 2009-09-16T17:18:19+00:00 |
| 3. ( <a href="#">0.5223</a> )                                                          | <a href="#">BIOMD0000000036</a> | Tyson1999_CircClock     | <a href="#">10545344</a> | 2008-08-21T11:51:20+00:00 |
| 4. ( <a href="#">0.4685</a> )                                                          | <a href="#">BIOMD0000000022</a> | Ueda2001_CircClock      | <a href="#">11403560</a> | 2008-08-21T11:45:52+00:00 |
| 5. ( <a href="#">0.4043</a> )                                                          | <a href="#">BIOMD0000000016</a> | Goldbeter1995_CircClock | <a href="#">8587874</a>  | 2008-08-21T11:43:17+00:00 |
| 6. ( <a href="#">0.3576</a> )                                                          | <a href="#">BIOMD0000000034</a> | Smolen2004_CircClock    | <a href="#">15111397</a> | 2008-08-21T11:50:44+00:00 |
| 7. ( <a href="#">0.2966</a> )                                                          | <a href="#">BIOMD0000000056</a> | Chen2004_CellCycle      | <a href="#">15169868</a> | 2009-09-17T13:54:43+00:00 |



A demo version showing the annotation based search can be found at:

<http://www.ebi.ac.uk/biomodels-demo/search>

<http://sombi.sourceforge.net/>

[Henkel et al., 2010] Henkel, R., Endler, L., Peters, A., Novère, N., and Waltemath, D. (2010). Ranked retrieval of computational biology models. *BMC Bioinformatics*, 11(1):423+

5 min

3 min

THE END

Thanks goes to:



Andre Peters



Nicolas Le Novère



Lukas Endler

many more...



**BioModels.Net**

