

Anni 2.1 Tutorial: Knowledge discovery

Version 1.0

Defining the starting concept

In this case, we will look for existing drugs that may be effective against a particular disease. We will follow the example described by Don Swanson, who discovered new therapies for Raynaud's disease using literature based knowledge discovery.

As a start, we will need to select the concept Raynaud's Disease, and import it in a concept set. To do this, we click the *import concepts* button as shown in Illustration 1.



Illustration 1: Click the import button to import the concept into a new concept set

Next, type “Raynaud's disease” in one of the empty cells, and click *map to concepts*. Anni will now look up the concept with that name. It will add a column to the table, specifying the preferred name of the concept. If you click on the concept, you will see the definition of the concept as shown in Illustration 2. Click *Ok* to close the import dialog.

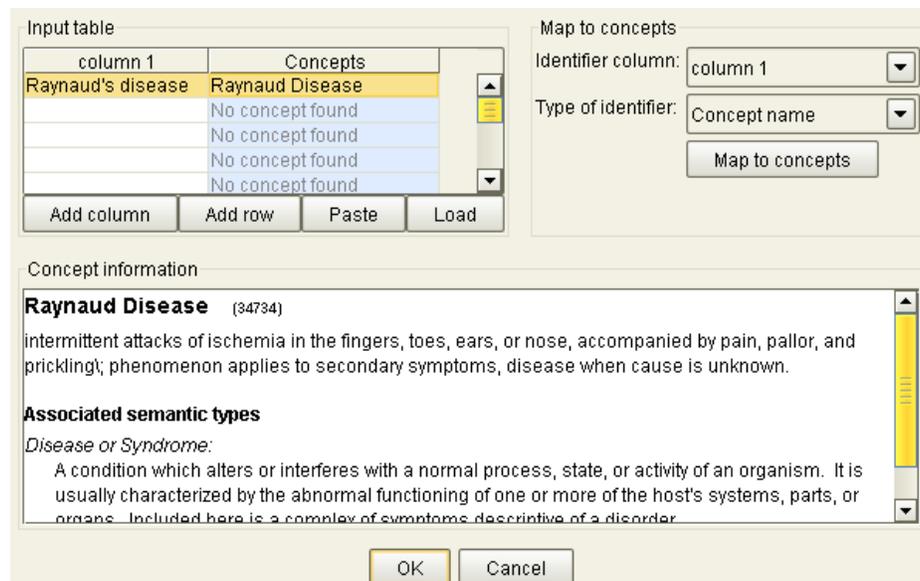


Illustration 2: Importing Raynaud's disease into a concept set

In the concept set explorer at the left side of the screen, you now see a concept set called “New concept

set” in the *User concept sets* folder. If you like, you can rename the concept to a more meaningful name. Right click on the concept set and select *Rename item* from the menu. Enter the new name, for instance “Raynaud disease”, and press enter.

Matching against drugs

Next we will match the concept profile of Raynaud's disease with the concept profiles of all the known pharmacological substances in Anni.

First, open the folder of predefined concept sets by clicking the  icon. Similarly, open the *Toxic Effect*, and select the concept set “Drug / Therapeutic Agent” as shown in Illustration 3.

Hold down the Ctrl button and click on the concept set you created in the previous part of this tutorial, containing the concept *Raynaud disease*. Now, both concept sets should be highlighted. Click the *Match Profiles to Profiles* button, as shown in Illustration 4.

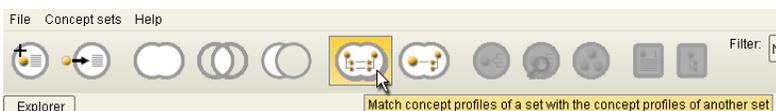
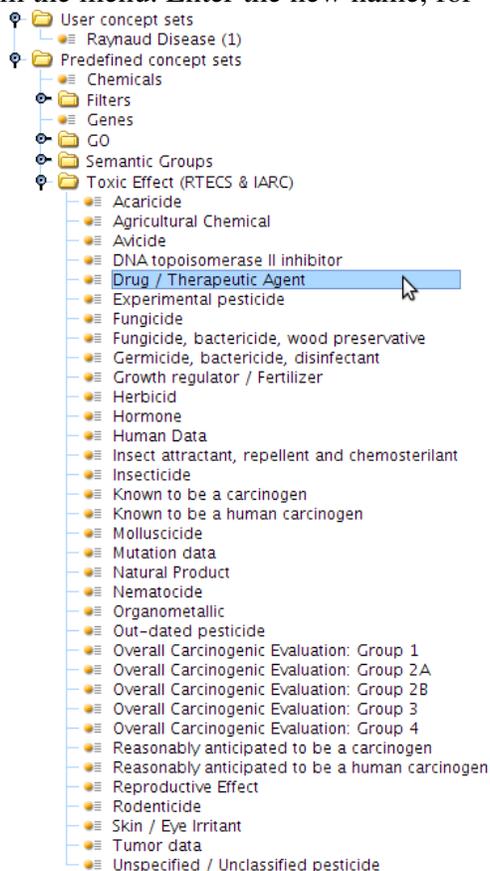


Illustration 4: After selecting both concept sets, click the *Match Profiles to Profiles* button

You will now see a dialog where you can specify which concept set will be displayed as rows, and which concept set will be displayed as columns in the resulting table. The biggest set, in this case “Drug / Therapeutic Agent” is best displayed as rows (A), and the smallest set is best displayed as columns (B), as shown in Illustration 5. Click *Ok* to close the dialog. Calculating the matching values will take some time.

Illustration 3: Selecting the predefined concept set 'Drug / Therapeutic Agent'

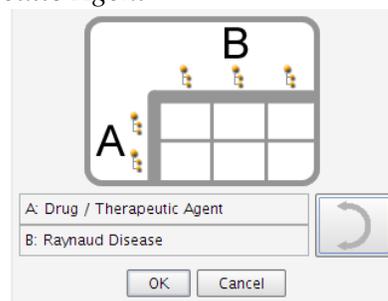


Illustration 5: Selecting the rows and columns for the matching table

The matching table

The matching table will show you the matching scores between the profiles of the concepts in the two sets that you selected. In this example, it shows the matching scores between Raynaud's disease on the one hand, and all drugs on the other. The matching score indicates the similarity between the two concept profiles. It is based on the number of concepts that the two profiles have in common, and the weights that those concepts have in the concept profiles. In other words: a high matching score indicates that the two concepts have many concepts in common. It is possible that concepts that have a high matching score have never been mentioned together. This could indicate new knowledge, and in this case could mean that there is drug that has never been reported to be applied to Raynaud's disease but could still be an effective treatment.

To find out which concepts have never been mentioned together with Raynaud's disease, click the

check box labeled *Highlight co-occurrence below* as shown in Illustration 6. The concepts with highlighted matching scores have never been mentioned together in the literature.

Concept	Sum	Raynaud Disease
Iloprost	0.0581	0.0581
inositol hexanicotinate	0.0522	0.0522
ketanserin	0.0342	0.0342
Dazoxiben	0.0339	0.0339
nifedipine	0.0334	0.0334
Sarpogrelate hydrochloride	0.0327	0.0327
prostaglandin I2	0.0325	0.0325
Cyclofenil	0.0308	0.0308
bosentan	0.0307	0.0307
Fendiline	0.0305	0.0305
bufloamedil	0.0301	0.0301
Moxislyte	0.0294	0.0294
Orphenadrine Citrate	0.0278	0.0278
prostaglandin E1	0.0261	0.0261
Beraprost sodium	0.0232	0.0232
limaprost	0.0227	0.0227
D-penicillamine	0.0224	0.0224
Moxislyte hydrochloride	0.022	0.022
Nicardipine	0.0212	0.0212
bleomycin A2	0.0207	0.0207

Illustration 6: The matching table, showing the matching scores between Raynaud's disease and drugs. Matching scores highlighted in light blue indicate no direct co-occurrence between the concepts exists.

In our example, the concept with the highest matching score and no direct co-occurrences is “Orphenadrine Citrate”. If we want to know why the matching score is so high, we can right-click on the matching score and select *Explain score*. Now we will see the annotation view.

The annotation view

The annotation view shows a breakdown of the matching score between the two concept profiles, as shown in Illustration 7.

Concept	ID	Contribution (%)	Orphenadrine Citrate	Raynaud Disease
Fingers	16129	25.846	0.0017	0.0058
Fingering	1832071	25.8426	0.0017	0.0058
Entire finger	1281584	25.8375	0.0017	0.0058
Digit structure	582802	5.2113	0.0007	0.0027
exposure to vibration	677519	1.811	0.0002	0.0036
Vibration – treatment	455941	1.6631	0.0002	0.0032
vibration – physical agent	459800	1.6631	0.0002	0.0032
Vibration	1883532	1.6631	0.0002	0.0032
carpal tunnel syndrome	7286	1.3411	0.0023	0.0002
Upper Extremity	1140618	1.1166	0.0013	0.0003

Illustration 7: Annotation of two concept profiles

The left column shows the concepts that the two profiles have in common. The third column, labeled

“Contribution %”, shows the percentage of the matching score that is due to that particular concept. The next two columns shows the weight of the concept in the concept profiles of the two main concepts.

In this example, a large part of the matching score is explained by the concepts fingers and exposure to vibration, and the first disease is “carpal tunnel syndrome”. If we want to know what the relationship is between carpal tunnel syndrome and Raynaud's disease, we can right-click on the cell on the intersection between the two concepts, and select *Find supporting documents*. This shows that there are 51 documents in Medline where both concepts have been mentioned together. The most recent 100 documents can be viewed by clicking on the link to PubMed.

Using a filter for matching

In the previous part of this tutorial, we matched concepts based on any other intermediate concept, including other diseases, genes, or even health care related organizations! If we want to restrict the matching, we will have to use a filter. A filter is a concept set containing those concepts that either should be used for matching (an “inclusive” filter), or those concepts that should not be used for matching (an “exclusive filter”).

Let's say for instance that we want to only match concepts based on physiological concepts. In the list of predefined concept sets, there is a folder called “Physiology”. However, this is a collection of concept sets, and we need to turn this into one big concept set. To do this, we need to select all Physiology concept sets. Click the first set, hold down the shift key and select the last concept set. Next we click on the *Merge concept sets* button as shown in Illustration 8. In the list of *User concept sets*, we now see a new concept set with a very long name. If you like, you can rename it to for instance “Physiology”.

Next, we right-click on this new set and select *Use as filter*. The name of the concept set will now also appear in the filter box in the toolbar. Click the radio button labeled *Include* to set the filter to “inclusive” as shown in Illustration 9.

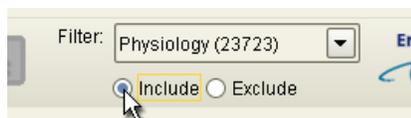


Illustration 9: Setting the filter to "inclusive"

From now on, the filter will apply to all operations in Anni. Repeat the steps described previously in this tutorial, starting at the section “Matching against drugs”, and notice the differences.

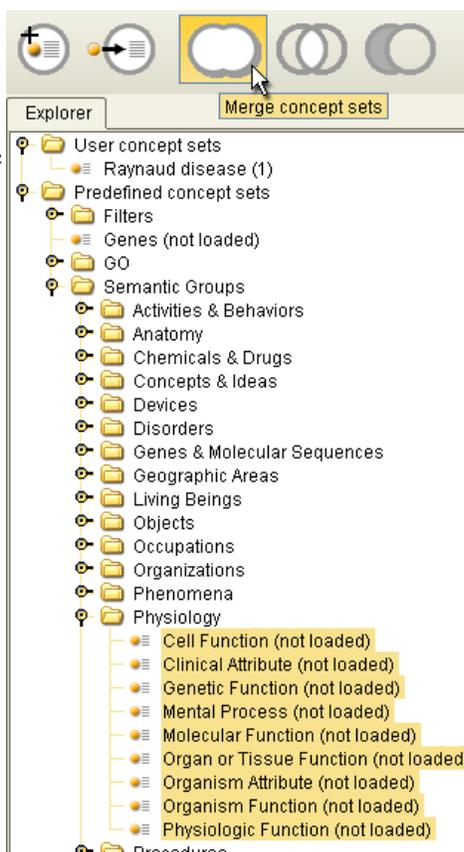


Illustration 8: Merge concept sets into a new set