Bisociative Exploration of Biological and Financial Literature Using Clustering

Oliver Schmidt[1], Janez Kranjc[2], Igor Mozetič[2], Paul Thompson[1], and Werner Dubitzky[1]

[1] University of Ulster, Northern Ireland, UK
schmidt-o1@email.ulster.ac.uk,
[2] Jozef Stefan Institute, Ljubljana, Slovenia

Abstract. The *bile acid and xenobiotic system* describes a biological network or system that facilitates detoxification and removal from the body of harmful xenobiotic and endobiotic compounds. While life scientists have developed a relatively comprehensive understanding of this system, many mechanistic details are yet to be discovered. Critical mechanisms are those which are likely to significantly further our understanding of the fundamental components and the interaction patterns that govern this systems gene expression and the identification of potential regulatory nodes. Our working assumption is that a creative information exploration of available bile acid and xenobiotic system information could support the development (and testing) of novel hypotheses about this system. To explore this we have set up an information space consisting of information from biology and finance, which we consider to be two semantically distant knowledge domains and therefore have a high potential for interesting bisociations. Using a cross-context clustering approach and outlier detection, we identify bisociations and evaluate their value in terms of their potential as novel biological hypotheses.

Keywords: Clustering, outlier detection, bisociative information exploration.

1 Introduction

Bisociative information exploration is based on the assumption that the pooling of information from different domains could facilitate the discovery of new knowledge. In this study we explore bisociative information discovery based on literature from molecular biology and finance. Our hypothesis is that the bisociative approach may help life scientists interested in the bile acid and xenobiotic system to generate (and possibly test) novel hypotheses which will ultimately support the discovery of biological mechanisms.

The presented approach is based on the work by Petrič et al. [10] who developed methods to investigate the role of outliers in literature-based knowledge discovery. Their approach rests upon the assumption that cluster outliers of two document sets with known classification can be used to discover new, useful

M.R. Berthold (Ed.): Bisociative Knowledge Discovery, LNAI 7250, pp. 438–451, 2012.

knowledge. In this context we define outliers as domain-labeled documents that are further away from the centroid of their knowledge domain then the majority of documents from its domain.

The work by Petrič et al.[10], which focuses on the domains of biology and medicine, differs from our approach in the way that we consider selected documents from two *unrelated* domains, namely, finance and biology. With *unrelated* domains we mean knowledge domains or domain theories (as defined in *Part I: Bisociation [2]*) that share less concepts than the knowledge domains of biology and medicine for instance. Therefore we expect to find less documents than between *related* domains, which in turn enables us to have a more detailed semi-automatic analysis of those documents.

We investigate the cluster outliers and their opposite-domain neighborhood in order to identify bisociations between biology and finance. In particular, we are looking for shared features in scientific abstracts across the two domains. Such features might be common terms or even sets of relationships within one domain which have correspondences in the other domain.

2 The Bile Acid and Xenobiotic System

The bile acid and xenobiotic system (BAXS) defines a biological network that facilitates two distinct but intimately overlapping physiological processes. The enterohepatic circulation and maintenance of bile acid concentrations (Fig. 1) and the detoxification and removal from the body of harmful xenobiotic (e.g. drugs, pesticides) and endobiotic compounds (e.g., steroid hormones) [8]. The system involves the coordination of several levels of gene activity, including control of mRNA and protein expression and regulation of metabolizing enzyme and transporter protein function in tissues such as liver, intestine/colon and kidney. Bile acids are necessary for the emulsification and absorption of dietary fats and are therefore valuable compounds, however as their build-up can cause harm, their concentrations need to be appropriately regulated and recycled. Similarly there is a requirement for a system that can "sense" the accumulation of xenobiotic and endobiotic compounds and facilitate their detoxification and removal from the body. The BAXS accomplishes this and maintains enterohepatic circulation (the circulation of biliary acids from the liver as depicted in Fig. 1) through a complex network of sensors in the form of nuclear receptors that function as ligand-activated transcription factors (see molecular interaction network depicted in Fig. 2). They serve to detect fluctuations in concentration of many compounds and initiate a physiological response by regulating the BAXS.

Transcriptional regulation by nuclear receptors[1] involves both activating and repressive effects upon specific "sets" of genes. There is considerable overlap exhibited between nuclear receptors in the genes they target and also the ligands

[1] Nuclear receptors are a class of proteins within the interior of cells responsible for sensing the presence of steroid and thyroid hormones and certain other molecules. In response, these receptors work in concert with other proteins to regulate the expression of specific genes, thereby controlling the development, homeostasis, and metabolism of the organism.

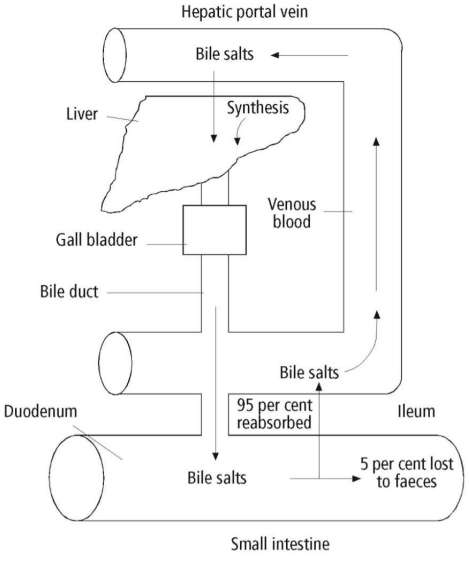

Fig. 1. The enterohepatic circulation system of the BAXS

that bind to and activate them, i.e. each gene has multiple functions within this system depending on the tissue it is expressed. It is these factors that contribute, for example, to the phenomenon of drug-drug interactions, e.g. between St. John's Wort and Cyclosporine or St. John's Wort and oral contraceptive [1,7].

The goal of the BAXS application within the BISON project is to support the discovery of hitherto unknown but important biological mechanisms in the BAXS. Critical mechanisms are those which are likely to significantly further our understanding of the fundamental components and the interaction patterns that govern BAXS gene expression and the identification of potential regulatory nodes. It has been established that the overall flux of the BAXS is achieved through a regulatory transcriptional network mediated through the activities of members of the nuclear receptors (such as FXR, LXR, PXR, CAR) and nuclear factors (such HNF1α, HNF4α). However, given the overall complexity of the bile acid/xenobiotic system it is difficult to assess the exact importance of each receptor and modulatory factor with respect to BAXS activity in different tissues. One of the key issues in the understanding of the BAXS is to decipher the components and the interaction patterns that govern BAXS gene expression and the identification of potential regulatory nodes. This understanding is essential to identify targets for treatment regimes, to understand the components impacting drug-drug interactions, to provide a framework for the design of large-scale, integrated prediction studies, and to aid in the definition of high-quality "gold standards" or research frameworks for future systems biology studies.

To investigate the potential of bisociative exploration of the BAXS, we are pooling two groups of information resources from the biological and financial domains respectively.

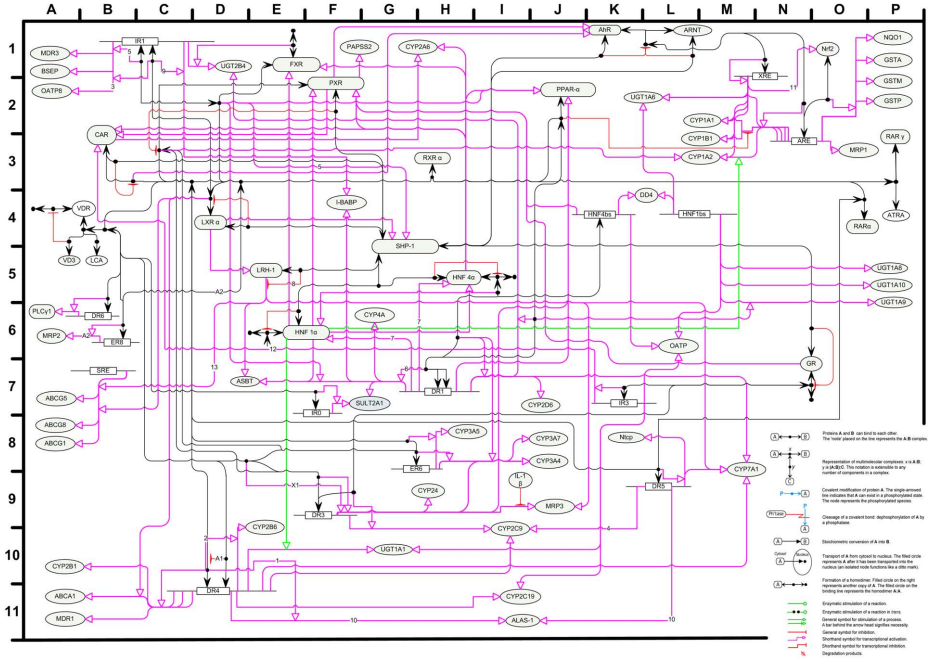

Fig. 2. The gene network and molecular interactions of the BAXS

3 Materials and Methods

OntoGen[2] is a semi-automatic and data-driven ontology editor facilitating the editing of topic ontologies (a hierarchy of topics connected with the subtopic_of relation). The editor uses text mining and machine learning techniques to help create ontologies from a set of text documents. For this the tool adopts k-means clustering and latent semantic indexing techniques(LSI) [6]. A screenshot of the tool is shown in Fig. 3.

In OntoGen each text document is represented as a term vector which is generated by the standard *bag of words* approach and the assignment of the *term frequency / inverse document frequency* ($TFIDF$)[3] measure to each term. The similarity between term vectors is calculated using Cosine similarity and shown in the center of Fig. 3. The similarity values are recalculated when the document selection changes. For further mentioning of the similarity between documents we refer to these similarity values calculated by OntoGen.

[2] `http://ontogen.ijs.si/`

[3] Elements of vectors are weighted with the TFIDF weights as follows [3]: the ith element of the vector containing frequency of the ith term is multiplied with $IDF_i = log(N/df_i)$, where N represents the total number of documents and df_i is document frequency of the ith term (i.e. the number of documents from the whole corpus in which the ith term appears).

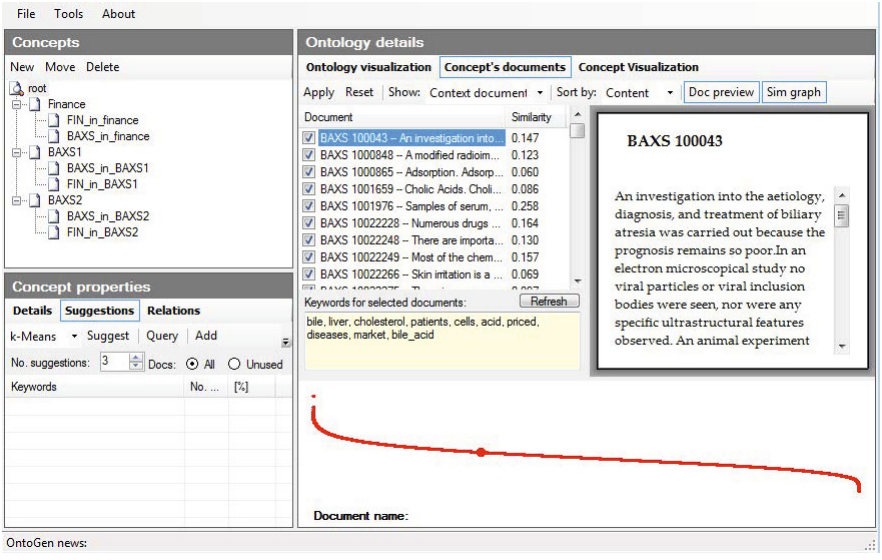

Fig. 3. Screenshot of the OntoGen's user interface. Left: The ontology concepts we created (top) and the functionality to create further sub concepts (bottom). Right: Details of the ontology, underlying documents and similarity graph information.

One of OntoGen's powerful features allows to visualize the similarity among a set of selected documents, which is called *Concept visualization*(as shown in Fig. 4). This visualization is created using dimensionality reduction techniques by combining linear subspace and multidimensional scaling methods. For a detailed description and explanation of OntoGen's main components and functionality we refer to work by Fortuna and colleagues [3,6,5,4].

In the following we describe how we generated and explored the document outliers across two domains: BAXS and finance. Petrič et al. [10] outline procedures that facilitate the identification of cross-context outliers with OntoGen. The main steps comprise the k-means clustering of documents with two different labels, the further subdivision of each cluster according to the labels and the outlier analysis of these misclassified documents in contrast to the clusters. Before going into the details of outlier detection, we describe the retrieval of the scientific documents for both domains. A *document* in this study refers to the *abstract* and associated *keywords* of a published scientific article.

The document corpus for the biological domain (as relevant to the BAXS) consists of documents from PubMed[4]. PubMed is a free resource containing over 20 million biomedical article citations and articles.

To compile a corpus of BAXS-relevant PubMed abstracts, we used keywords and phrases that reflect important concepts in relation to BAXS research. Furthermore, we restricted the search to articles that discuss these concepts in the context of human biology (ignoring other species). We used the following

[4] `www.pubmed.gov`

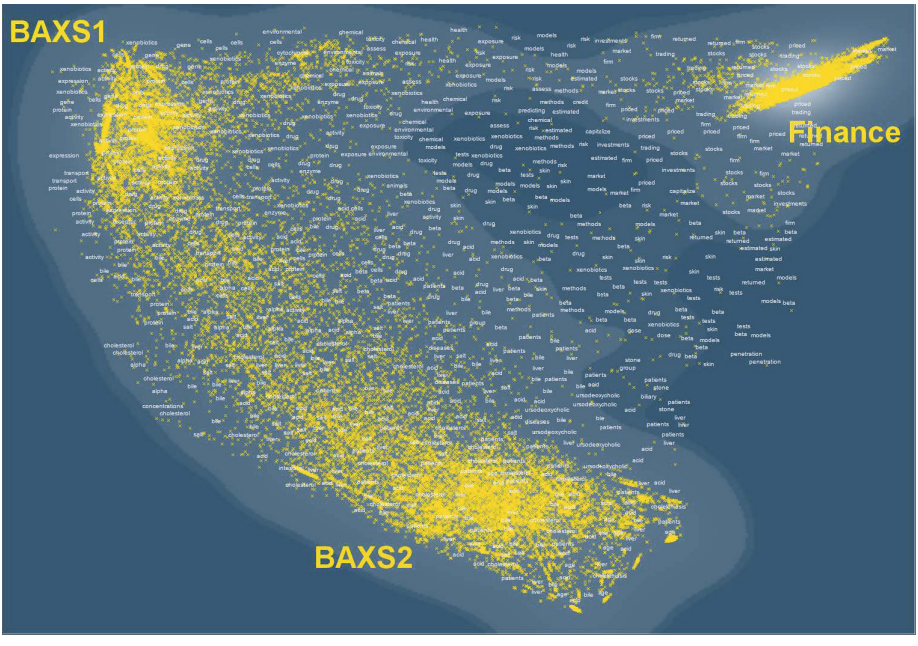

Fig. 4. OntoGen's concept visualization of all documents. Yellow crosses denote documents and the white labels depict document terms. The 3 potential clusters are labeled accordingly

PubMed query to select the articles:

```
(‘‘bile acids and salts’’ [MeSH Terms] OR
(‘‘bile’’ [All Fields] AND ‘‘acids’’ [All Fields] AND
‘‘salts’’ [All Fields] ) OR ‘‘bile acids and salts’’
[All Fields] OR (‘‘bile’’ [All Fields] AND
‘‘acids’’ [All Fields]) OR ‘‘bile acids’’ [All Fields]) OR
(‘‘xenobiotics’’ [MeSH Terms] OR
‘‘xenobiotics’’ [All Fields] ) AND ‘‘humans’’ [MeSH Terms]
```

The query resulted in 21 565 articles of which 16 106 had an abstract. In addition to the abstracts, we retrieved all articles with the MeSH terms provided by PubMed, i.e., we included also articles with MeSH[5] terms only. With this approach we compiled 21 276 documents containing either abstracts, MeSH terms or both.

The information resources from the financial domain are abstracts from the financial literature. We obtained these from the Journal Storage[6] (JSTOR). Currently, JSTOR contains approximately 1224 journals, which are categorized

[5] Medical Subject Headings (MeSH) provides a vocabulary of ca. 25 000 terms used to characterize the content of biomedical document such as articles in scientific journals.

[6] `http://dfr.jstor.org/`

into 19 collections and 53 disciplines. The archive currently offers approximately 295 000 individual journal issues and about 6.4 million articles of which about 3.2 million articles are full-text articles available online. Out of 27 613 available finance articles that we retrieved, 7674 provided an abstract. In addition to the abstracts, we retrieved a set of JSTOR-generated keywords for each abstract.

We created a simple text file for the BAXS and finance articles, respectively, each file containing all the abstracts in that domain. Within each file the documents were organized as a sequence of lines, each line representing a document, and the first word in the line is used as the title of the document. A list of 532 English stop words was used to filter out the "low-content" words followed by Porter stemming to reduce inflected (or derived) words to their stem [9]. For the construction of the similarity graph the document labels were ignored.

After some experimentation and visual analysis of the clusters, we generated the final set of clusters using k-means clustering with $k = 3$. Viewing the resulting clustering of the document vectors (as shown in Fig. 4), one can easily recognize the three clusters. In order to get an idea about the information content of the clusters and the area between them, a more detailed view on Fig. 4 (shown in Fig. 5) is provided.

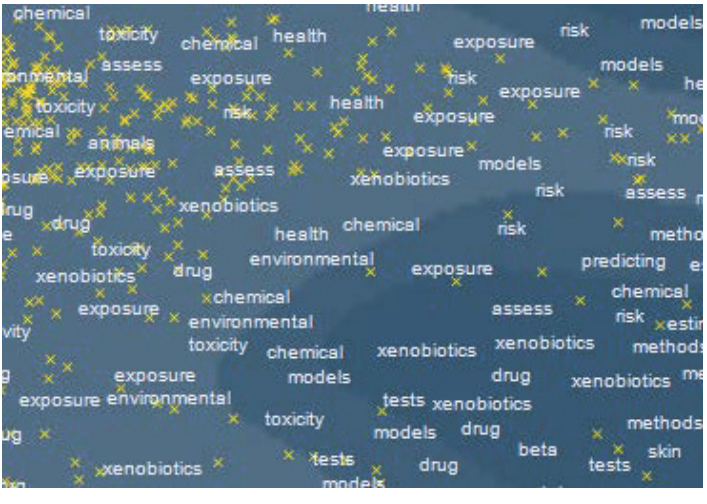

Fig. 5. Zoomed in detail from the top centre of Figure 4 (area between ***BAXS1*** and ***Finance cluster***).

Another interesting feature of OntoGen is its ability to determine the most common keywords for each cluster. OntoGen uses two methods to extract the keywords (1) using centroid vectors to get descriptive keywords and (2) Support Vector Machine classification to extract distinctive keywords. In this study we considered keywords extracted from method (1) only. The keywords provide users with an idea of the meaning of the clusters. According to this we characterized the three clusters as follows:

1. The ***Finance Cluster*** characterized by the keywords *priced*, *market*, *stocks*, *returned*, *firm*,*trading*, *investments*, *models*, *portfolio* and *rate*.
2. The ***BAXS1 Cluster*** with the keywords *cells*, *activity*, *protein*, *drug*, *transport*, *receptor*, *enzyme*, *xenobiotics*, *expression* and *gene*.
3. The ***BAXS2 Cluster*** with the keywords *bile*, *cholesterol*, *patients*, *liver*, *acid*, *diseases*, *age*, bile acid, *biliary* and *ursodeoxycholic*.

In order to determine the outliers for each cluster, we separated the documents of each of the three clusters according to the assigned document labels `BAXS` and `FIN`. The resulting topic ontology is depicted in Fig. 6.

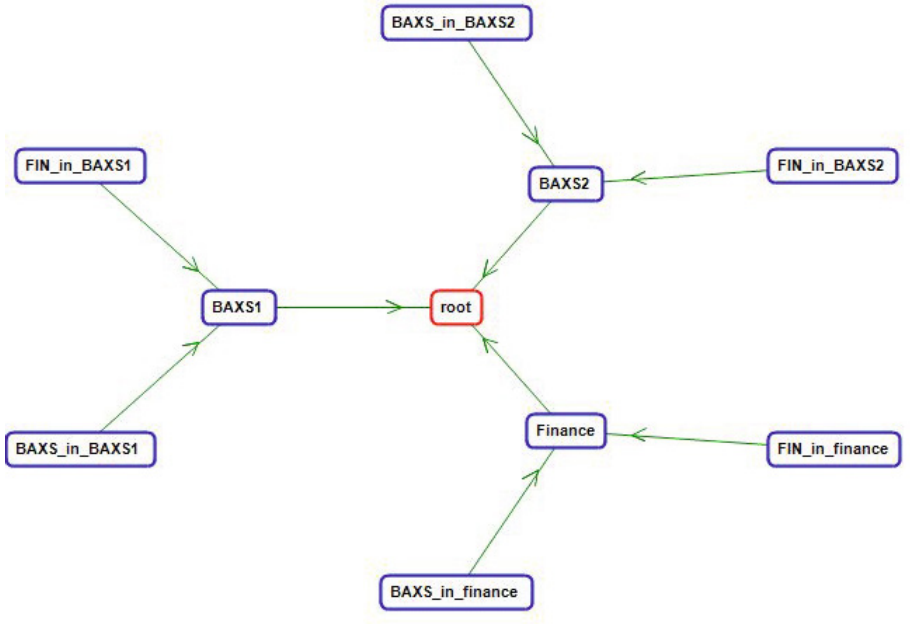

Fig. 6. Visualization of the topic ontology with the root in the middle and its three ascending nodes/clusters (*Finance*, *BAXS1*, *BAXS2*) of which each has two ascending nodes/clusters corresponding to either labels `BAXS` or `FIN`.

Fig. 6 suggests that each cluster contains outliers, i.e., *BAXS_in_Finance* for `BAXS`-labeled documents in the Finance cluster, *FIN_in_BAXS1* for `FIN`-labeled documents in the BAXS1 cluster, and *FIN_in_BAXS2* for `FIN`-labeled documents in the BAXS2 cluster. A detailed breakdown of the distribution of outliers over the clusters is provided in Table 1 as a contingency table.

The rows in the table described the number of documents falling into one of the three clusters, and the columns describe the number of documents labeled `BAXS` and `FIN` respectively. Out of 7674 `FIN`-labeled documents, 7671 were assigned to cluster Finance, 2 documents were assigned to cluster BAXS1, and

Table 1. Contingency table: Overview of document distribution over the clusters *Finance*, *BAXS1* and *BAXS2*

Cluster/Label	FIN	BAXS	**Total**
Finance	7671	59	7730
BAXS1	2	9129	9131
BAXS2	1	12 088	12 089
Total	7674	21 276	28 950

1 document was assigned to cluster BAXS2. Out of 21 276 BAXS-labeled documents, 59 were assigned to cluster Finance, 9129 documents were assigned to cluster BAXS1, and 12 088 documents were assigned to cluster BAXS2. From a total of 28 950 documents 26,5% are labeled as FIN and 73,5% as BAXS. The number of documents shared between the clusters is 26,7% in the Finance cluster, 31,5% in BAXS1 cluster and 41,8% in BAXS2 cluster of all documents.

In order to show the ratio between the documents with *initial label* and the *cluster membership* of documents, we combined the documents and outliers from the two clusters BAXS1 and BAXS2 to a single cluster, BAXS, as shown in the confusion matrix in Table 2. Out of 7674 FIN documents, 3 were assigned to combined BAXS cluster, and out of 21 276 BAXS documents, 59 were assigned to the Finance cluster. That is, from all FIN documents 0.04% were outliers and from all BAXS documents 0.28% were outliers.

Table 2. Confusion matrix: Overview of correctly and misclassified documents for the labels FIN and BAXS

		Initial label	
		FIN	BAXS
Cluster membership	**Finance**	7671	59
	BAXS	3	21 217

The aim of this study is to interpret the outliers in context to the documents which are most similar for each cluster. As looking at 62 outliers and their most similar neighbours for each cluster would be manually not feasible we reduced the amount of BAXS outliers within the Finance cluster.

From the 59 BAXS outliers within the Finance cluster, we decided to consider 13 of them to be relevant for this study. The decision making process was accomplished by experts analysing the 59 BAXS documents and filtering them. We selected the outliers that were most relevant to BAXS but also most promising to find relationships in finance. The list of outliers and the topic covered is shown in Table 3.

As next step we investigated the most similar documents for each outlier document for each cluster. We considered the 5 most similar documents for the BAXS clusters (BAXS1 and BAXS2) and the 6 most similar documents for the Finance cluster. The reason for considering one more document from the Finance cluster was that FIN-labeled documents were generally much shorter than BAX

Table 3. General topic of outlier documents

Outlier	Topic
BO_01	Data analysis using statistical models
BO_02	Study about how paracetamol dissolves in different conditions
BO_03	A new model in pharmacokinetics is introduced
BO_04	Paper discusses legal criteria and judicial precedents related to hormesis
BO_05	How to calculate reference doses for toxic substances
BO_06	Nonlinear system to model concentration of hormones important for menstrual cycle
BO_07	Paper about volume of distribution and mean residence time of pharmaceutics in the body
BO_08	Book about chronic kidney disease and future perspective for Japan
BO_09	Application of data mining methods for biomonitoring and usage of xenobiotics
BO_10	Xenobiotics in the air close to industrial centres affect mechanisms of inheritance
BO_11	Data analysis about colorectal polyp prevention
BO_12	Statistical data analysis for Alzheimer's disease
BO_13	Costs and effectiveness of anti-fungal agents for patients are assessed
FO_01	Analysis about how new drug launches affect life expectancy
FO_02	Analysis of Russian's investment in transport infrastructure
FO_03	Relationship between choice of treatment for illness and getting a job

documents, and were therefore considered to offer less information than the BAXS documents. The set of abstracts we retrieved contained on average 8 sentences for PubMed abstracts and 4 for JSTOR.

As OntoGen wasn't designed or intended to be used in such a way we adopted the following procedure to obtain the documents that are most similar to an outlier. First, we selected the outlier of interest and deselected all other documents. Then, using OntoGen, we recalculated the similarity of all documents for this outlier. The most similar documents to the outlier are then listed below the outlier.

In order to find the most similar documents from a different cluster, one had to assign the considered outlier to the other cluster. This was achieved by the selection of the outlier and selecting OntoGen'S *move* function (shown in the top left corner in Fig. 7). After the outlier was moved to one of the other clusters, the procedure was repeated to obtain the most similar neighborhood documents.

Petrič et al. [10] considered only neighborhood documents within the same cluster for the interpretation of outliers. Our approach extends this approach by taking into account the neighborhood documents from the other clusters. Thus, possible relationships between the clusters can be assessed, where the outliers serve as link between the most similar neighborhood documents. Table 4 lists the common bridging terms between the outliers and their neighborhood documents that we found.

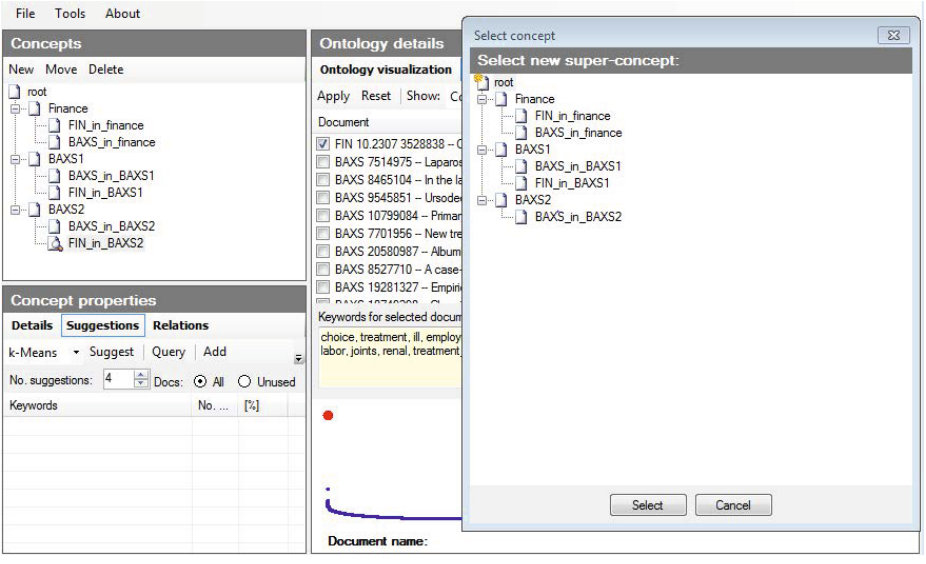

Fig. 7. Screenshot showing how the outlier document FIN 10.2307 3528838 in cluster *BAXS2* can be moved to different clusters

4 Results and Discussion

Initially, we wanted to look at all 16 outliers and how they might be linked or related to the topic they cover. For this we first looked at the selected BAXS outlier documents and how they cover the topics of all `BAXS` outliers within the Finance cluster. As shown in Fig. 8, the 13 outliers (the dark labels) seem to cover most of the topic space of all 59 outliers. We realized that the outliers are not evenly distributed over this space, which is due to the bias created by manual selection by the experts. In fact, there are more outliers close to each other on the left side of the diagram then on the right side.

If we look at the topics covered by each outlier (as shown in Table 3), we see that most of them cover topics related to clinical studies, statistical analysis or models related to BAXS in one way or another.

The `FIN` outliers, on the other hand, do not seem to be similar to each other. There do not seem to be any obvious relationships between the life expectancy affected by drug launches (FO 01), how Russians invest in their transport infrastructure (FO 02), or how the choice of treatment for illness is related to getting a job (FO 03).

Therefore, we analyzed the outliers in detail and determined the terms they share with their most similar neighbor documents. The results of this analysis are summarized in Table 4. The table lists the most frequent bridging terms (b-terms) between each outlier and their neighbors.

Then we looked at the bridging terms for the BAXS outliers within each cluster. The b-terms between the 13 BAXS outliers and the documents in the

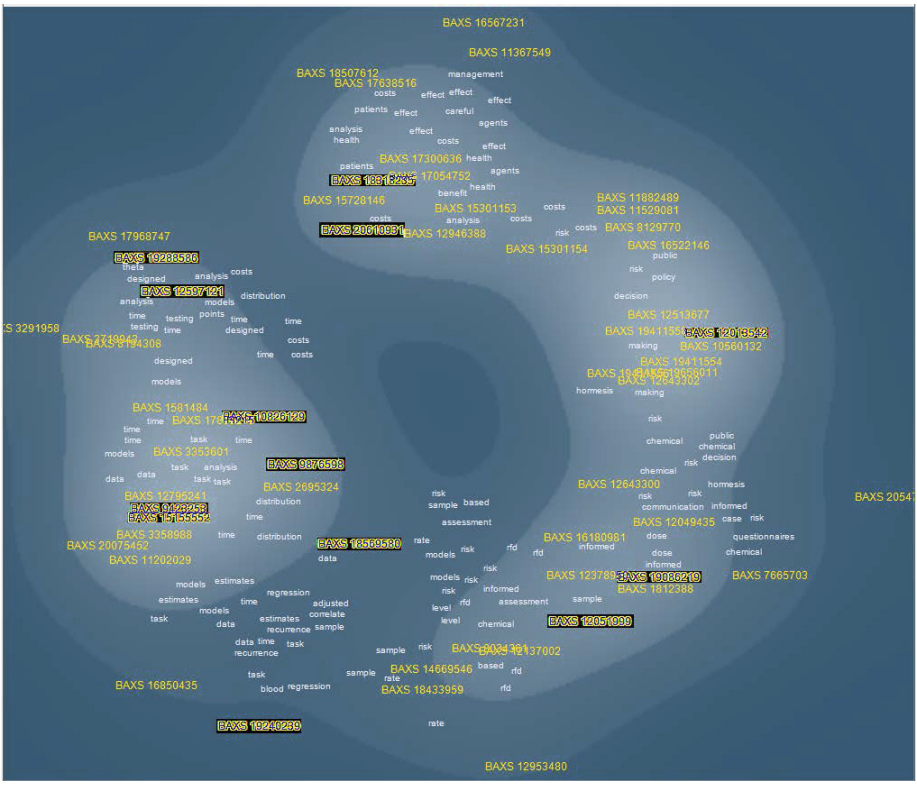

Fig. 8. Concept visualization of all `BAXS` outlier documents within the Finance cluster. The relevant outliers under investigation are highlighted in black

Finance cluster are usually general terms such as *model, event, predict, estimate or risk*. The b-terms within the *BAXS1* cluster and *BAXS2* appear to be more specific, i.e., biological or medical terms such as *hygiene, hormone, pharmacokinetic, colon cancer*. The b-terms between the 3 *Finance* outliers within *BAXS1* and *BAXS2* do not provide any new insights compared to Table 3

Meaningful bisociative relationships between the outliers and their neighbor documents could not be found. As this neglects the neighbor documents within the other clusters, we decided to search for relationships between the neighbor documents from the clusters. For this we picked the outliers BO 05 and FO 02 (see Table 4) due to their common b-terms in each cluster.

The BO 05 outlier relates finance models concerned with the risk of investment to abstracts which analyze the risk of cancer depending on different factors. The most promising outlier for this was FO 02, as it relates BAXS abstracts to transporter pathways within and across cells with the transporting infrastructure within and across countries.

Table 4. Discovered bridging terms via outliers with most similar neighbours from the clusters *Finance*, *BAXS1* and *BAXS2*.

Outlier	Finance	BAXS1	BAXS2
BO 01	event, model, simulation	model	regression
BO 02	model, predict, statistical	model, dissolution	dissolution
BO 03	curve, distribution, long-term, model	model, pharmacokinetic	model, pharmacokinetic (PK), kinetic
BO 04	decision-making, regulatory, agency	hormesis, toxic, regulator	humans (Mesh)
BO 05	risk, estimate, observe	NOAEL, BMD	risk
BO 06	cycles, non-linear systems, model	hormone	cycle, women, hormone
BO 07	volume, estimate	clearance, estimate, pharmacokinetic	volume
BO 08	japanese, assume, analyse	filtration	humans, middle aged (Mesh)
BO 09	decision-tree, predict, model	environmental	PCB, biomonitoring, HCB
BO 10	air, risk	hygiene, air, xenobiotics (Mesh)	xenobiotics
BO 11	estimation, method	colorectal, colon cancer	recurrence, prevention, polyp
BO 12	model, predict	model, analysis, PLSDA	Alzheimer's disease, patient
BO 13	cost, hospitals	antifungal	cost, leukemia
FO 01	health, expenditure, cost-effectiveness	drug, database	drug
FO 02	Russia, transport, transit	transport	transit, transport
FO 03	choice	renal	treatment, disease, chronic, cholestasis

5 Conclusions

In this case study we investigated the potential bisociations between the finance and BAXS domain based on document outliers as determined by a cross-context clustering approach.

One main issue in this study is the asymmetric nature of knowledge, i.e., we have more knowledge about the BAXS than about the finance domain. Another issue is the asymmetric nature of the data sources (73% `BAXS` documents and 27% `Finance` documents). Both put a strong bias on the discovered outliers in this study and therefore reduce the quality of the results. Based on this study it would appear that the most promising method to find potential bisociations is to look at the neighbor documents from the different clusters related to one outlier. The use of scientific abstracts only could be another reason for the lack of finding interesting relationships between the domains. More work is needed to explore

this approach to bisociative information discovery but the approach presented here shows promise in the discovery of novel connections between domains

References

1. Barone, G., Gurley, B., Ketel, B., Lightfoot, M., Abul-Ezz, S.: Drug interaction between St. John's wort and cyclosporine. The Annals of Pharmacotherapy 34(9), 1013–1016 (2000)
2. Dubitzky, W., Kötter, T., Schmidt, O., Berthold, M.R.: Towards Creative Information Exploration Based on Koestler's Concept of Bisociation. In: Berthold, M.R. (ed.) Bisociative Knowledge Discovery. LNCS (LNAI), vol. 7250, pp. 11–32. Springer, Heidelberg (2012)
3. Fortuna, B., Grobelnik, M., Mladenič, D.: Visualization of text document corpus (2005)
4. Fortuna, B., Grobelnik, M., Mladenič, D.: Semi-automatic data-driven ontology construction system (2006)
5. Fortuna, B., Grobelnik, M., Mladenic, D.: OntoGen: Semi-automatic Ontology Editor. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007, Part II. LNCS, vol. 4558, pp. 309–318. Springer, Heidelberg (2007)
6. Fortuna, B., Mladenič, D., Grobelnik, M.: Semi-automatic Construction of Topic Ontologies. In: Ackermann, M., Berendt, B., Grobelnik, M., Hotho, A., Mladenič, D., Semeraro, G., Spiliopoulou, M., Stumme, G., Svátek, V., van Someren, M. (eds.) EWMF 2005 and KDO 2005. LNCS (LNAI), vol. 4289, pp. 121–131. Springer, Heidelberg (2006)
7. Hall, S.D., Wang, Z., Huang, S., Hamman, M.A., Vasavada, N., Adigun, A.Q., Hilligoss, J.K., Miller, M., Gorski, J.C.: The interaction between St. John's wort and an oral contraceptive[ast]. Clinical Pharmacology and Therapeutics 74(6), 525–535 (2003)
8. Kliewer, S.A.: The nuclear pregnane X receptor regulates xenobiotic detoxification. The Journal of Nutrition 133(7), 2444S–2447S (2003)
9. Natarajan, J., Berrar, D., Hack, C.J., Dubitzky, W.: Knowledge discovery in biology and biotechnology texts: A review of techniques, evaluation strategies, and applications. Critical Reviews in Biotechnology 25(1-2), 31–52 (2005)
10. Petrič, I., Cestnik, B., Lavrač, N., Urbančič, T.: Outlier detection in cross-context link discovery for creative literature mining. The Computer Journal (2010), doi:10.1093/comjnl/bxq074