Thalia
User manual

Axel J. Soto, Piotr Przybyła, Sophia Ananiadou
National Centre for Text Mining (NaCTeM), UK

Introduction

Thalia (**T**ext mining for **H**ighlighting, **A**ggregating and **L**inking **I**nformation in **A**rticles) is a semantic search engine that can recognise concepts occurring in biomedical abstracts indexed on PubMed (https://www.ncbi.nlm.nih.gov/pubmed/). It currently recognises eight types of concepts, namely: chemicals, diseases, drugs, genes, metabolites, proteins, species and anatomical entities.

This user manual describes the main components and options of the web-based tool as well as it demonstrates the semantic capabilities of Thalia.

The webpage for this project can be found on http://nactem.ac.uk/Thalia (password protected while under review—u: “reviewer”, p: “bi”), where further information about the interface as well as any updates on Thalia are posted.

Main interface

Thalia can be accessed using a web browser[1]. Once loaded, after showing a welcome dialog, which also links to this manual, Thalia prompts the user with the initial search interface as shown in Figure 1.

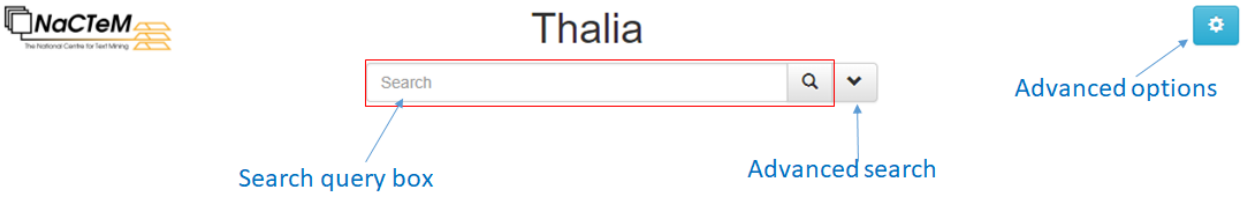


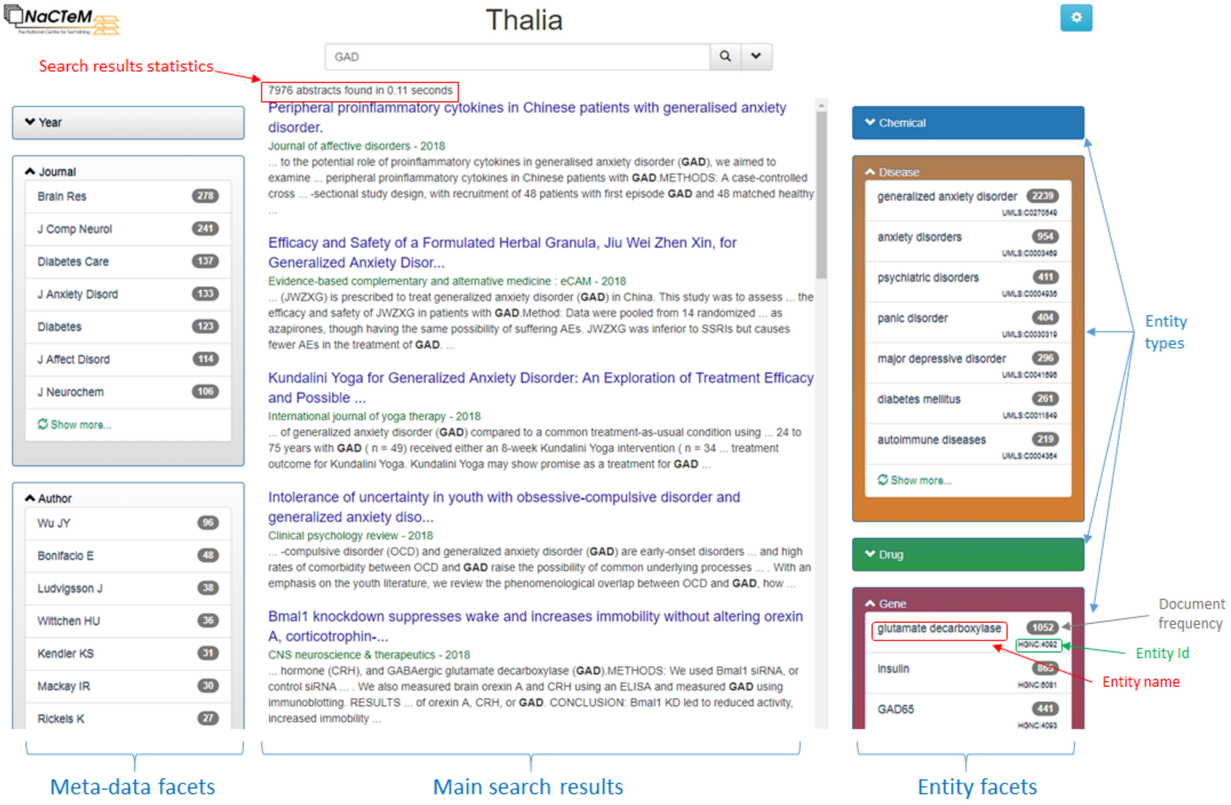
Figure 1: Initial search interface

A text-based search can be executed by typing a query into the search query box and hitting “Enter” key or clicking the magnifying glass icon. For example, if the user enters the text string *GAD*, Thalia will return a list of abstracts from PubMed containing the word *GAD* as shown in

[1] Google Chrome, Safari and Firefox have been tested to be compatible with Thalia.

Figure 2. The middle panel shows a list of the first 20 search results, including the total number of matching documents and the time taken, which are shown at the top. The user can scroll down the middle panel to inspect each of the snippets and click on the bottom arrows to show the next set of 20 results, if needed.

Side panels show two expandable sets of result facets. The facets on the left are related to publication meta-data—i.e. *Publication Year*, *Journal Name*, *Author*, *Publication Type* and *MeSH tags*—while the ones on the right correspond to entity facets—i.e. *Chemical*, *Disease*, *Drug*, *Gene*, *Metabolite*, *Protein*, *Species* and *Anatomical*—. Within each facet there is a list of values with a number indicating the document frequency as a badge. This means that after clicking on a specific item, the next search will be filtered by the documents containing that facet value, and the total number of filtered documents will be the number indicated in the badge. In the case of entities, the facets also show the corresponding concept identifier in an ontology associated with the entity type. Note that the search query box can also parse more complex queries, including phrasal queries—e.g. *"homo sapiens"* (note the quotes)—or boolean queries—e.g. *GAD AND diabetes*.



7976 abstracts found in 0.11 seconds

Peripheral proinflammatory cytokines in Chinese patients with generalised anxiety disorder.
Journal of affective disorders - 2018
... to the potential role of proinflammatory cytokines in generalised anxiety disorder (**GAD**), we aimed to examine ... peripheral proinflammatory cytokines in Chinese patients with **GAD**.METHODS: A case-controlled cross ... -sectional study design, with recruitment of 48 patients with first episode **GAD** and 48 matched healthy ...

Efficacy and Safety of a Formulated Herbal Granula, Jiu Wei Zhen Xin, for Generalized Anxiety Disor...
Evidence-based complementary and alternative medicine : eCAM - 2018
... (JWZXG) is prescribed to treat generalized anxiety disorder (**GAD**) in China. This study was to assess ... the efficacy and safety of JWZXG in patients with **GAD**.Method: Data were pooled from 14 randomized ... as azapirones, though having the same possibility of suffering AEs. JWZXG was inferior to SSRIs but causes fewer AEs in the treatment of **GAD**. ...

Kundalini Yoga for Generalized Anxiety Disorder: An Exploration of Treatment Efficacy and Possible ...
International journal of yoga therapy - 2018
... of generalized anxiety disorder (**GAD**) compared to a common treatment-as-usual condition using ... 24 to 75 years with **GAD** (n = 49) received either an 8-week Kundalini Yoga intervention (n = 34 ... treatment outcome for Kundalini Yoga. Kundalini Yoga may show promise as a treatment for **GAD** ...

Intolerance of uncertainty in youth with obsessive-compulsive disorder and generalized anxiety diso...
Clinical psychology review - 2018
... -compulsive disorder (OCD) and generalized anxiety disorder (**GAD**) are early-onset disorders ... and high rates of comorbidity between OCD and **GAD** raise the possibility of common underlying processes ... . With an emphasis on the youth literature, we review the phenomenological overlap between OCD and **GAD**, how ...

Bmal1 knockdown suppresses wake and increases immobility without altering orexin A, corticotrophin-...
CNS neuroscience & therapeutics - 2018
... hormone (CRH), and GABAergic glutamate decarboxylase (**GAD**).METHODS: We used Bmal1 siRNA, or control siRNA ... . We also measured brain orexin A and CRH using an ELISA and measured **GAD** using immunoblotting. RESULTS ... of orexin A, CRH, or **GAD**. CONCLUSION: Bmal1 KD led to reduced activity, increased immobility ...

Journal	
Brain Res	278
J Comp Neurol	241
Diabetes Care	137
J Anxiety Disord	133
Diabetes	123
J Affect Disord	114
J Neurochem	106

Author	
Wu JY	96
Bonifacio E	48
Ludvigsson J	38
Wittchen HU	36
Kendler KS	31
Mackay IR	30
Rickels K	27

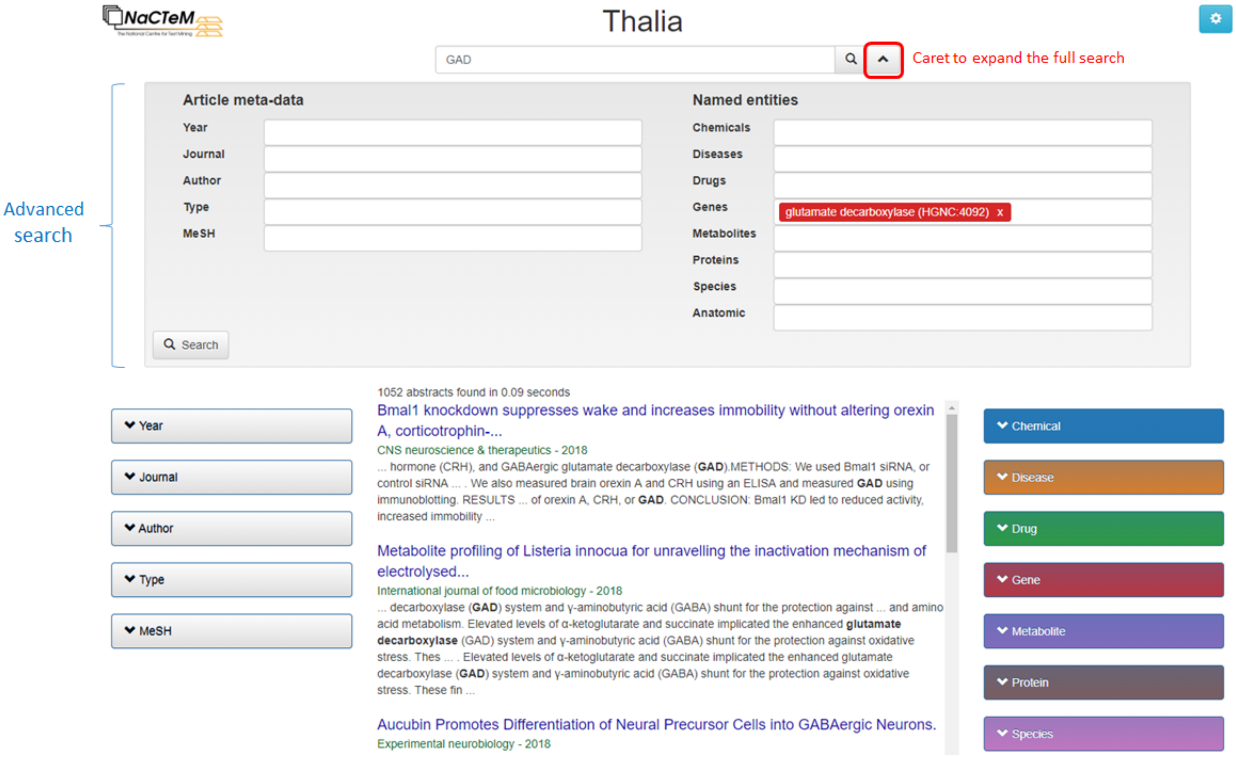
Disease		
generalized anxiety disorder	2239	UMLS:C0270549
anxiety disorders	954	UMLS:C0003469
psychiatric disorders	411	UMLS:C0004936
panic disorder	404	UMLS:C0030319
major depressive disorder	296	UMLS:C0041696
diabetes mellitus	261	UMLS:C0011849
autoimmune diseases	219	UMLS:C0004364

Gene		
glutamate decarboxylase	1052	HGNC:4092
insulin	865	HGNC:6081
GAD65	441	HGNC:4093

Figure 2: Results for a text-based search

We can observe from the concepts listed in the entity facets in Figure 2 that *GAD* is an ambiguous term that in certain contexts can refer to the disease *generalized anxiety disorder* or the gene *glutamate decarboxylase*. This is where the semantic capacity of Thalia can be used to disambiguate which of the two senses were intended by the user who formulated the query.

Let us assume that the user is interested in the gene. After clicking on *glutamate decarboxylase* from the gene facet, a new search is triggered that enforce the HGNC:4092 concept to be present in the documents retrieved. Although this concept has *glutamate decarboxylase* as its most common name, it has other possible names, such as: *GAD*, *GAD1* and *glutamate decarboxylase 1*, *glutamate decarboxylase 1 (brain, 67kD)* and *glutamate decarboxylase 1 (brain, 67kDa)*. The result of this procedure is shown in Figure 3, where the number of results retrieved is 1052 (down from 7976)[2]. At the same time we can see that the clicked concept has been added to the advanced search panel. This advanced search panel can be visualised by clicking on the caret besides the query search box.



GAD

Caret to expand the full search

Advanced search

Genes: glutamate decarboxylase (HGNC:4092) x

1052 abstracts found in 0.09 seconds

Bmal1 knockdown suppresses wake and increases immobility without altering orexin A, corticotrophin-...
CNS neuroscience & therapeutics - 2018
... hormone (CRH), and GABAergic glutamate decarboxylase (**GAD**).METHODS: We used Bmal1 siRNA, or control siRNA ... . We also measured brain orexin A and CRH using an ELISA and measured **GAD** using immunoblotting. RESULTS ... of orexin A, CRH, or **GAD**. CONCLUSION: Bmal1 KD led to reduced activity, increased immobility ...

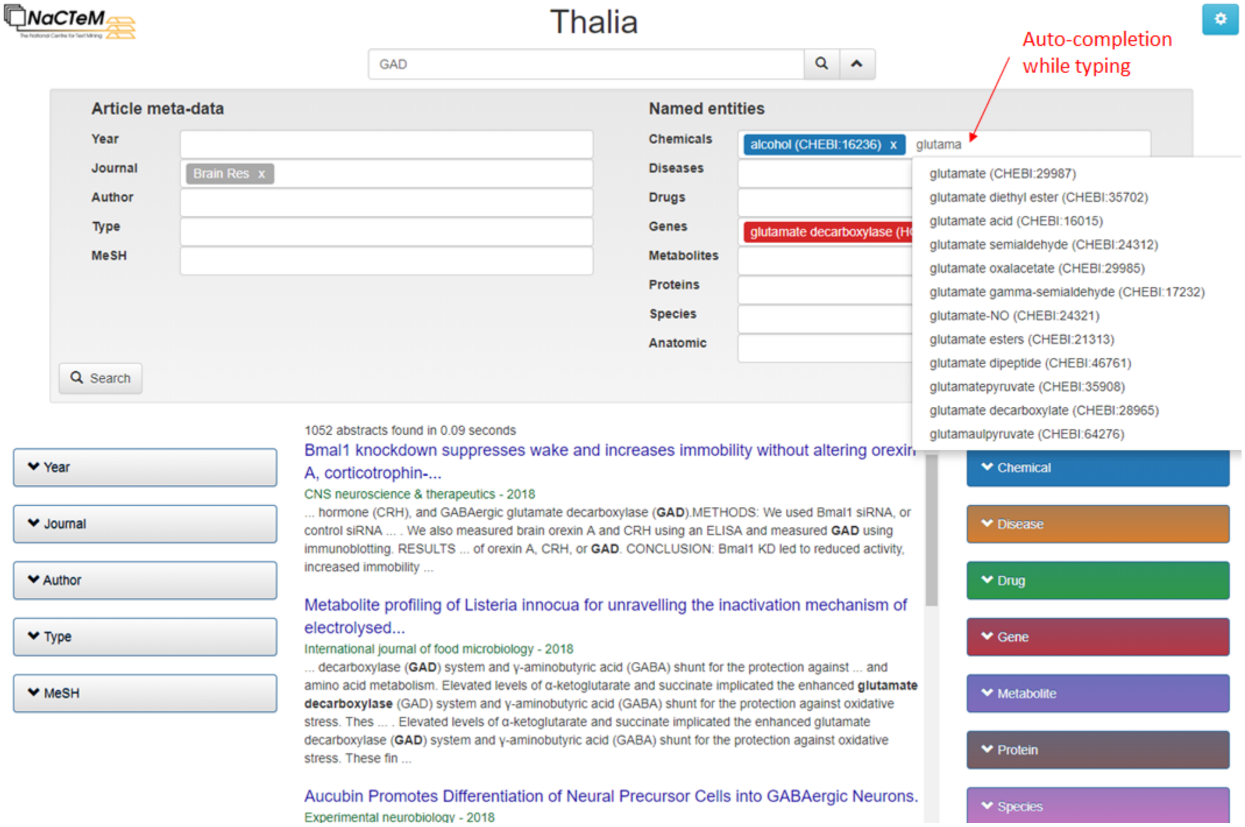
Metabolite profiling of Listeria innocua for unravelling the inactivation mechanism of electrolysed...
International journal of food microbiology - 2018
... decarboxylase (**GAD**) system and γ-aminobutyric acid (GABA) shunt for the protection against ... and amino acid metabolism. Elevated levels of α-ketoglutarate and succinate implicated the enhanced **glutamate decarboxylase** (GAD) system and γ-aminobutyric acid (GABA) shunt for the protection against oxidative stress. Thes ... . Elevated levels of α-ketoglutarate and succinate implicated the enhanced glutamate decarboxylase (**GAD**) system and γ-aminobutyric acid (GABA) shunt for the protection against oxidative stress. These fin ...

Aucubin Promotes Differentiation of Neural Precursor Cells into GABAergic Neurons.
Experimental neurobiology - 2018

Figure 3: Thalia after clicking on one value of the gene facet

[2] The exact numbers may be different due to the continuous update carried out on Thalia.

The advanced search panel allows to type in article meta-data as well as named entities to include in the search query. To facilitate the accurate specification of the terms in the advanced search, an auto-completion mechanism is in place, as it is shown in Figure 4.



Auto-completion while typing

GAD

Journal: Brain Res x

Chemicals: alcohol (CHEBI:16236) x glutama

Genes: glutamate decarboxylase (H

glutamate (CHEBI:29987)
glutamate diethyl ester (CHEBI:35702)
glutamate acid (CHEBI:16015)
glutamate semialdehyde (CHEBI:24312)
glutamate oxalacetate (CHEBI:29985)
glutamate gamma-semialdehyde (CHEBI:17232)
glutamate-NO (CHEBI:24321)
glutamate esters (CHEBI:21313)
glutamate dipeptide (CHEBI:46761)
glutamatepyruvate (CHEBI:35908)
glutamate decarboxylate (CHEBI:28965)
glutamaulpyruvate (CHEBI:64276)

1052 abstracts found in 0.09 seconds

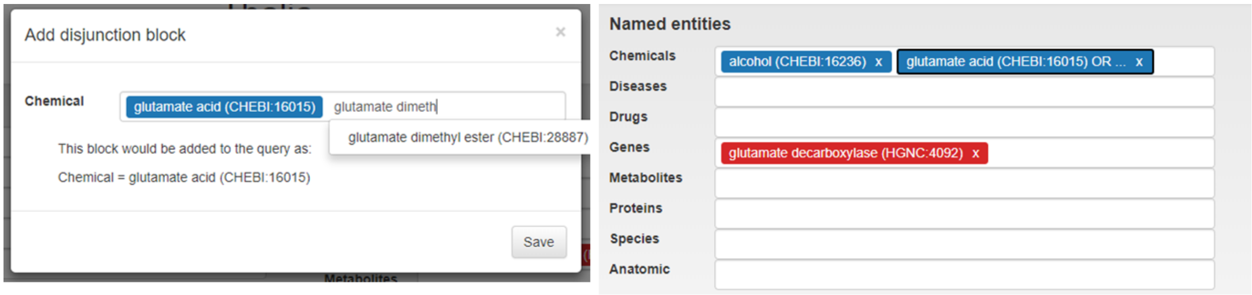
Bmal1 knockdown suppresses wake and increases immobility without altering orexin A, corticotrophin-...
CNS neuroscience & therapeutics - 2018
... hormone (CRH), and GABAergic glutamate decarboxylase (**GAD**).METHODS: We used Bmal1 siRNA, or control siRNA ... . We also measured brain orexin A and CRH using an ELISA and measured **GAD** using immunoblotting. RESULTS ... of orexin A, CRH, or **GAD**. CONCLUSION: Bmal1 KD led to reduced activity, increased immobility ...

Metabolite profiling of Listeria innocua for unravelling the inactivation mechanism of electrolysed...
International journal of food microbiology - 2018
... decarboxylase (**GAD**) system and γ-aminobutyric acid (GABA) shunt for the protection against ... and amino acid metabolism. Elevated levels of α-ketoglutarate and succinate implicated the enhanced **glutamate decarboxylase** (GAD) system and γ-aminobutyric acid (GABA) shunt for the protection against oxidative stress. Thes ... . Elevated levels of α-ketoglutarate and succinate implicated the enhanced glutamate decarboxylase (**GAD**) system and γ-aminobutyric acid (GABA) shunt for the protection against oxidative stress. These fin ...

Aucubin Promotes Differentiation of Neural Precursor Cells into GABAergic Neurons.
Experimental neurobiology - 2018

Figure 4: Auto-completion of values in the advanced search panel

Once a value is entered in the advanced search—either typed-in or selected from the facets—, the term is shown enclosed with a tag. In the case of named entities, if normalised, the tag will exhibit the concept identifier in parentheses. As each entity type has a color associated to it, the tag will be color-coded accordingly. In the case of the meta-data values, the tags are always gray. Each of these tags will be interpreted in the query as a conjunction, which means that a boolean AND operation is used to aggregate all the values. In order to use a disjunction, the user has to click on the tag, which will lead to a new window for entering the concepts for disjunction (Figure 5a). Once the desired disjunction is entered, this type of tags is shown using a dark thick border (Figure 5b).



Add disjunction block

Chemical: glutamate acid (CHEBI:16015) glutamate dimeth

glutamate dimethyl ester (CHEBI:28887)

This block would be added to the query as:

Chemical = glutamate acid (CHEBI:16015)

Named entities

Chemicals: alcohol (CHEBI:16236) x; glutamate acid (CHEBI:16015) OR ... x

Diseases

Drugs

Genes: glutamate decarboxylase (HGNC:4092) x

Metabolites

Proteins

Species

Anatomic

a) b)

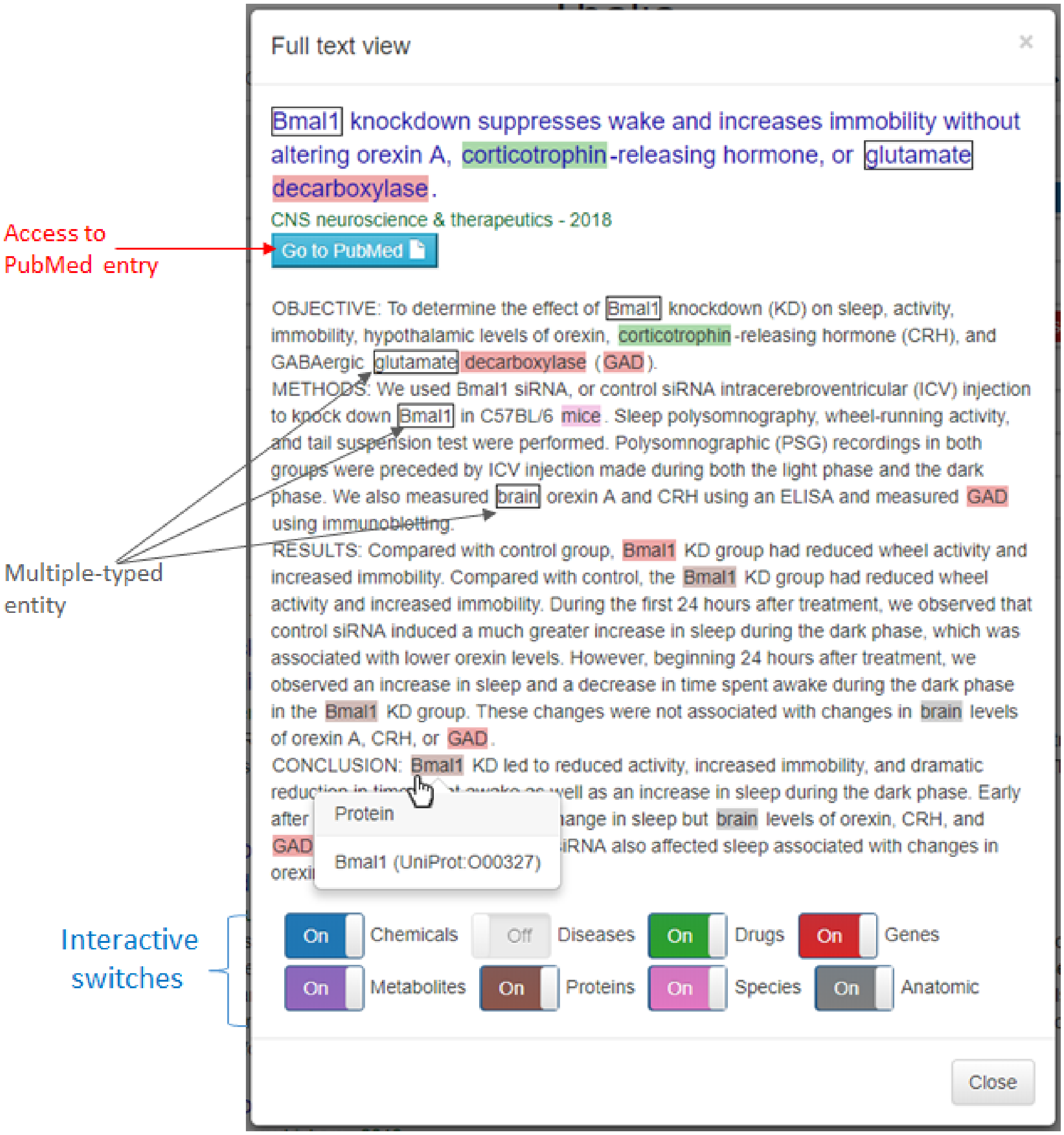
Figure 5: Process of adding a term to be used as a disjunction in the query. a) Entering a term to be aggregated using a disjunction. b) The entities that form a disjunction block are shown with a thicker border.

An entity normalised to a concept in an ontology can be inspected by Ctrl-clicking (Command-clicking for Mac users) on the tag, which will open the web page of the ontology concept being clicked. All the ontologies are openly available, except for UMLS (used for diseases), which requires free user registration.

Full text view

Retrieved abstracts can be inspected upon clicking on an entry of the retrieved document list. Thalia will show a window similar to the one in Figure 6. Each recognised entity is highlighted with the color of its corresponding type. Multiple-typed entities are visualised with a black-bordered rectangle. At the bottom of the panel, there are toggle buttons that allow switching on or off the highlighting of entities of a specific type. At the top of the panel there is a light blue button to open a new page with the PubMed entry of that article.

Hovering over a highlighted entity will pop up a label with its concept identifier. When an entity is clicked, the tool will open the ontology web page for the corresponding concept. In the case of entities highlighted as multiple-typed, the user needs to switch off some entity types until just one type is highlighted for the entity, before the user can be redirected to the ontology entry.



Full text view

Bmal1 knockdown suppresses wake and increases immobility without altering orexin A, corticotrophin-releasing hormone, or glutamate decarboxylase.

CNS neuroscience & therapeutics - 2018

Go to PubMed

OBJECTIVE: To determine the effect of Bmal1 knockdown (KD) on sleep, activity, immobility, hypothalamic levels of orexin, corticotrophin-releasing hormone (CRH), and GABAergic glutamate decarboxylase (GAD). METHODS: We used Bmal1 siRNA, or control siRNA intracerebroventricular (ICV) injection to knock down Bmal1 in C57BL/6 mice. Sleep polysomnography, wheel-running activity, and tail suspension test were performed. Polysomnographic (PSG) recordings in both groups were preceded by ICV injection made during both the light phase and the dark phase. We also measured brain orexin A and CRH using an ELISA and measured GAD using immunoblotting. RESULTS: Compared with control group, Bmal1 KD group had reduced wheel activity and increased immobility. Compared with control, the Bmal1 KD group had reduced wheel activity and increased immobility. During the first 24 hours after treatment, we observed that control siRNA induced a much greater increase in sleep during the dark phase, which was associated with lower orexin levels. However, beginning 24 hours after treatment, we observed an increase in sleep and a decrease in time spent awake during the dark phase in the Bmal1 KD group. These changes were not associated with changes in brain levels of orexin A, CRH, or GAD. CONCLUSION: Bmal1 KD led to reduced activity, increased immobility, and dramatic reduction in tim[…]at awake as well as an increase in sleep during the dark phase. Early after […]ange in sleep but brain levels of orexin, CRH, and GAD […]iRNA also affected sleep associated with changes in orexi[…]

Protein

Bmal1 (UniProt:O00327)

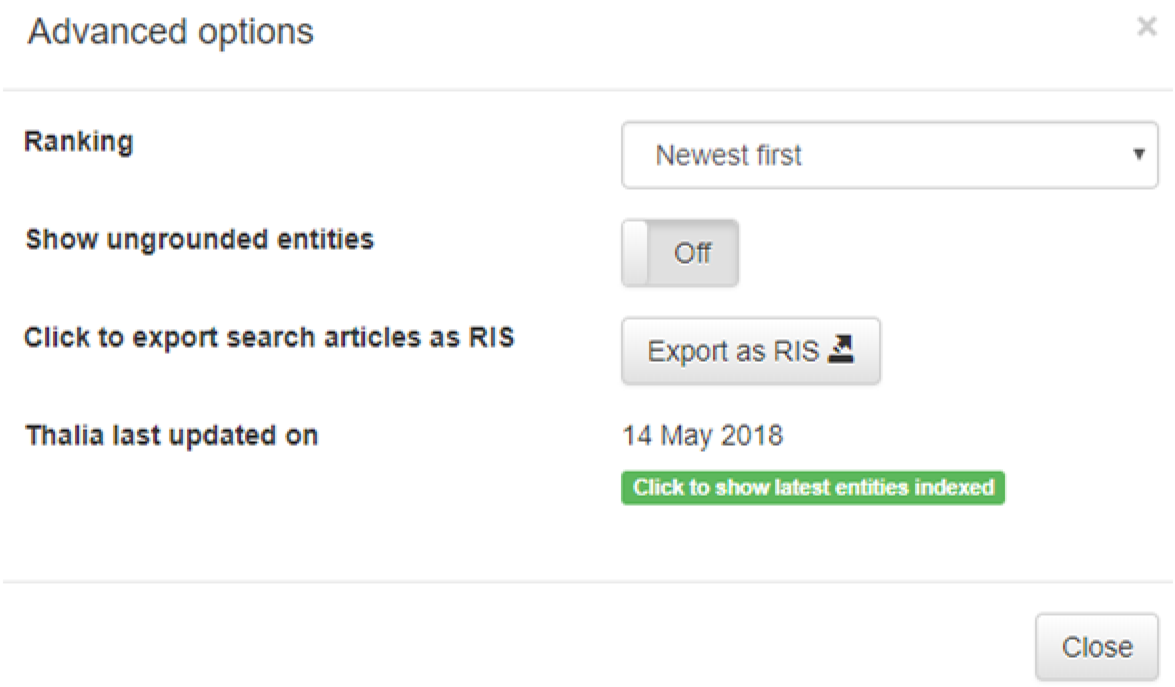
- On Chemicals
- Off Diseases
- On Drugs
- On Genes
- On Metabolites
- On Proteins
- On Species
- On Anatomic

Close

Figure 6: Full-text view

Advanced options

Advanced options for Thalia can be accessed by clicking the light blue gear icon located at the top right of the screen (see Figure 1). There are four main items accessible from this panel as it is shown in Figure 7.



Advanced options

Ranking: Newest first

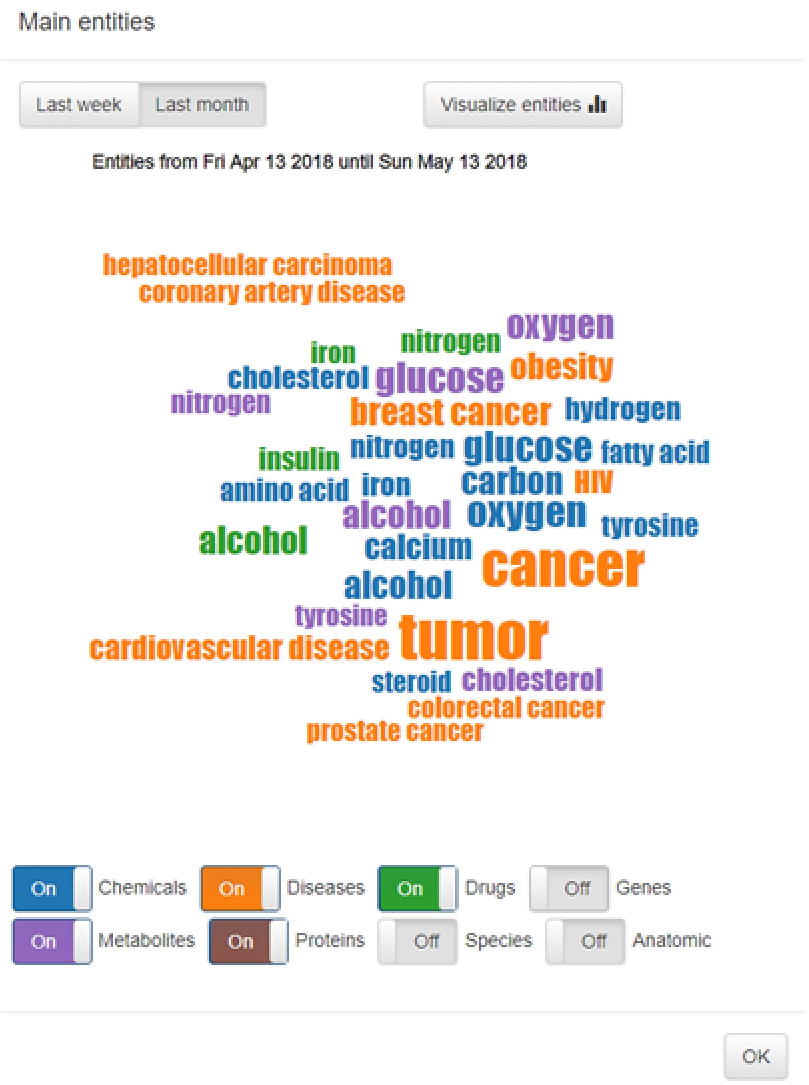
Show ungrounded entities: Off

Click to export search articles as RIS

Thalia last updated on: 14 May 2018

Figure 7: Advanced options

The first option allows to choose ranking retrieved documents either chronologically or by relevance. For generic queries it is usually most convenient to focus on the latest results first. If the query gets more specific, it may be better to rank by relevance. The second option determines whether ungrounded entities are taken into account or not. By default, only entities that could be normalised to concepts are shown to the user (during autocompletion, as facetted values or as highlights in the document view). The user can, however, take into account entities that failed to be grounded by enabling this second option. The third option allows to export the set of retrieved documents as a RIS file (there is a limit currently set to 1000 for the maximum number of documents that can be exported at once). Finally, there is a date that shows when Thalia was last updated. The user can click on this item to show which were the latest entities updated in the form of a word cloud (Figure 8). After choosing between *Last week* or *Last month* and clicking on the *Visualize entities* button, the user can inspect the most frequent entities in the documents added during the chosen period. Similarly to the Full Text View, the user can interact with the switches to turn on or off certain types of entities.



Entities from Fri Apr 13 2018 until Sun May 13 2018

On Chemicals

On Diseases

On Drugs

Off Genes

On Metabolites

On Proteins

Off Species

Off Anatomic

Figure 8: Word cloud with the most frequent entities indexed by Thalia during the last month

Attribution

If you use Thalia API in your research or project, please cite the following article.

Axel J. Soto, Piotr Przybyła, Sophia Ananiadou, “Thalia: Semantic search engine for biomedical Abstracts”, Bioinformatics, Under review.