

Google Scholar Indexing for Institutional Repositories

Presentation for Couperin consortium

October 2021

Monica Westin, Google Scholar

Overview

1. How Google Scholar indexes repositories
2. Collaboration among HAL and Google Scholar
3. The crucial role of accurate metadata for indexing
4. How to ensure best possible metadata on the HAL platform
5. How to ensure best possible metadata on the DSpace platform
6. How to check your repository's coverage in Scholar
7. Additional guidelines and resources
8. Questions and discussion

How Google Scholar indexes repositories

How the Google Scholar indexing system works

- The Scholar indexing system crawls the entire web looking for scholarly publications: articles, books, reports, theses, conference proceedings, preprints ...
- The indexing system identifies scholarly content, determines each item's bibliographic metadata, and groups all versions of an item together with this metadata in search results

The screenshot shows a Google Scholar search interface. At the top, the Google Scholar logo is on the left, and a search bar contains the text "Directed complete poset models of T1 spaces" with a magnifying glass icon on the right. Below the search bar, it indicates "Articles" and "2 results (0.03 sec)". On the left side, there are filters for "Any time", "Since 2019", "Since 2018", "Since 2015", and "Custom range...". The main search result is titled "Directed complete poset models of T_1 spaces" and is from a PDF at nie.edu.sg. The author is D Zhao, X Xi, and the publication is in the Proceedings of the Cambridge Philosophical Society, 2018. The abstract describes a poset model of a topological space X and states that every T_1 space has a (bounded complete algebraic) poset model. At the bottom of the result, there are icons for a star, a document, and a link, followed by the text "Cited by 1 Related articles All 4 versions". The text "All 4 versions" is circled in red.

Google Scholar

"Directed complete poset models of T1 spaces"

Articles 2 results (0.03 sec)

Any time
Since 2019
Since 2018
Since 2015
Custom range...

Directed complete poset models of T_1 spaces [PDF] nie.edu.sg

D Zhao, X Xi - ... Proceedings of the Cambridge Philosophical Society, 2018 - cambridge.org

A poset model of a topological space X is a poset P such that the subspace $\text{Max}(P)$ of the Scott space ΣP is homeomorphic to X , where $\text{Max}(P)$ is the set of all maximal points of P . Every T_1 space has a (bounded complete algebraic) poset model. It was, however, not ...

☆ 📄 Cited by 1 Related articles All 4 versions

We try to make repositories visible worldwide

- For uniquely held items (e.g. dissertations), repository content is the primary link

Primary link →

[THE MITOCHONDRIAL NATURE OF THE DNA REPAIR PROTEIN APE1](#)

[A Barchiesi - 2018 - iris.sissa.it](#)

Mitochondria arose around two billion years ago from the engulfment of an α -proteobacterium by a precursor of the modern eukaryotic cell 1. Nowadays they are the principal source of energy production and metabolism that are fundamental processes for the survival and wellbeing of the cell. Mitochondria are organelles with a double membrane, the outer (OM) and the inner membrane (IM), and together they delimit two aqueous compartments: the matrix, in the inner side, and the intermembrane space (IMS)(Figure 1).

☆ [🔗](#) [Related articles](#) [All 3 versions](#) [🔗](#)

Showing the best result for this search. [See all results](#)

[\[PDF\] sissa.it](#)

← Access link

- For formally published articles, repository content appears as the access link and/or in “All XX versions”

Directed complete poset models of T_1 spaces

[D Zhao, X Xi - ... Proceedings of the Cambridge Philosophical Society, 2018 - cambridge.org](#)

A poset model of a topological space X is a poset P such that the subspace $\text{Max}(P)$ of the Scott space ΣP is homeomorphic to X , where $\text{Max}(P)$ is the set of all maximal points of P . Every T_1 space has a (bounded complete algebraic) poset model. It was, however, not known whether every T_1 space has a directed complete poset model and whether every sober T_1 space has a directed complete poset model whose Scott topology is sober. In this paper we give a positive answer to each of these two problems. For each T_1 space X , we ...

☆ ⓘ Cited by 1 Related articles [All 4 versions](#)

[\[PDF\] nie.edu.sg](#)



All versions

Directed complete poset models of T_1 spaces

[D Zhao, X Xi - ...](#) Proceedings of the Cambridge Philosophical Society, 2018 - cambridge.org

A poset model of a topological space X is a poset P such that the subspace $\text{Max}(P)$ of the Scott space ΣP is homeomorphic to X , where $\text{Max}(P)$ is the set of all maximal points of P . Every T_1 space has a (bounded complete algebraic) poset model. It was, however, not ...

☆ [Cited by 1](#) [Related articles](#)

← publisher version

Directed complete poset models of T_1 spaces

[D Zhao, X Xi - 2018 - repository.nie.edu.sg](#)

A poset model of a topological space X is a poset P such that the subspace $\text{Max}(P)$ of the Scott space P is homeomorphic to X , where $\text{Max}(P)$ is the set of all maximal points of P . Every T_1 space has a (bounded complete algebraic) poset model. It was, however, not ...

[Cited by 1](#)

← repository version

[CITATION] Directed complete poset models of T_1 spaces

[D Zhao, X Xi - Mathematical Proceedings of the Cambridge ..., 2018 - adsabs.harvard.edu](#)

[Cited by 1](#)

← citation

Directed complete poset models of T_1 spaces

[D ZHAO, X XI - Mathematical Proceedings of the Cambridge ..., 2018 - search.proquest.com](#)

A poset model of a topological space X is a poset P such that the subspace $\text{Max}(P)$ of the Scott space [...] P is homeomorphic to X , where $\text{Max}(P)$ is the set of all maximal points of P . Every T_1 space has a (bounded complete algebraic) poset model. It was, however, not ...

[Cited by 1](#)

← aggregator version

What Scholar needs for indexing repository publications

- Access to crawl the site
- Way to find all urls for articles-- usually sitemap for repositories
- Bibliographic information in the form of machine-readable metadata tags (“metatags”), on by default for DSpace repositories past version 1.7

Bibliographic metatags tell the Scholar indexing system what the metadata for an article is: title, author, publication date, etc.



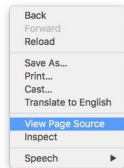
“Citation_pdf_url” metatag tells the indexing system which file to associate with this metadata



```
<meta name="citation_title" content="The testis isoform of the phosphorylase kinase catalytic subunit (PhK-T) plays a critical role in regulation of glycogen mobilization in developing lung">
<meta name="citation_author" content="Liu, Li">
<meta name="citation_author" content="Rannels, Stephen R.">
<meta name="citation_author" content="Falconieri, Mary">
<meta name="citation_author" content="Phillips, Karen S.">
<meta name="citation_author" content="Wolpert, Ellen B.">
<meta name="citation_author" content="Weaver, Timothy E.">
<meta name="citation_publication_date" content="1996/05/17">
<meta name="citation_journal_title" content="Journal of Biological Chemistry">
<meta name="citation_volume" content="271">
<meta name="citation_issue" content="20">
<meta name="citation_firstpage" content="11761">
<meta name="citation_lastpage" content="11766">
<meta name="citation_pdf_url"
content="http://www.example.com/content/271/20/11761.full.pdf">
```


View source code from repository item page to view metatags

Right click or keyboard command to “View Page Source,” depending on your browser



Search HTML source for “citation_” to view metatags



Search Repository

Repositório do ISPA / Psicologia da Saúde / PSAU - Artigos em revistas nacionais

Please use this identifier to cite or link to this item: <http://hdl.handle.net/10400.12/1089>

Title: Adaptação e estudo da escala de valor da saúde

Author: Pimenta, Filipa
Leal, Isabel Pereira
Maroco, João

Keywords: Valor
Saúde
Fumadores
Ex-fumadores
Escala
Value
Health
Smokers
Ex-smokers
Scale

Issue Date: 2009

Publisher: Sociedade Portuguesa de Psicologia da Saúde

Citation: Psicologia, Saúde & Doenças, 10 (2), 217-225

Abstract: O presente estudo pretende averiguar se pessoas com diferentes comportamentos de consumo de tabaco valorizam a sua saúde da forma discrepante; é igualmente objectivo desta investigação adaptar a Escala de Valor da Saúde (Lau, Hartman, & Ware, 1986) à população Portuguesa. Para tal aplicou-se o instrumento a uma amostra de 380 estudantes universitários (fumadores regulares, ocasionais e exfumadores). Os resultados demonstram que pessoas com diferentes padrões de consumo tabágico atribuem um valor elevado à sua saúde, não existindo diferenças significativas entre os três grupos (F=1,594; p=0,205). O estudo das características psicométricas do instrumento, já utilizado em estudos anteriores, salienta que, numa amostra de 380 alunos do ensino superior, esta escala não é adequada para medir a variável em questão. ---- ABSTRACT ---- The aim of the present study is to analyse if people with diverse smoking patterns value their health differently. It is also an objective of this research to adapt the Health Value Scale (Lau, Hartman, & Ware, 1986) to the Portuguese population. A sample of 380 college students (regular and occasional smokers, and ex-smokers) has filled this instrument. The results show that people with different smoking patterns attribute a high value to their health and that there are no significant differences between the three groups (F=1,594; p=0,205). The study of the psychometric characteristics of the scale, which has been used in previous studies, evidence that, in a sample of 380 college students, this instrument is not adequate to measure the variable health value.

view-source: <http://repositorio.ispa.pt/handle/10400.12/1089>

```
<meta name="citation_author" content="Pimenta, Filipa" />
<meta name="citation_author" content="Leal, Isabel Pereira" />
<meta name="citation_author" content="Maroco, João" />
<meta name="citation_issn" content="1645-0086" />
<meta name="citation_journal_title" content="Psicologia, Saúde & Doenças" />
<meta name="citation_date" content="2009" />
<meta name="citation_abstract_html_url" content="https://repositorio.ispa.pt/handle/10400.12/1089" />
<meta name="citation_lastpage" content="225" />
<meta name="citation_language" content="por" />
<meta name="citation_pdf_url" content="https://repositorio.ispa.pt/bitstream/10400.12/1089/1/PSD%202009%2010%282%29%20217-225.pdf" />
<meta name="citation_keywords" content="Valor; Saúde; Fumadores; Ex-fumadores; Escala; Value; Health; Smokers; Ex-smokers; Scale; article" />
<meta name="citation_publisher" content="Sociedade Portuguesa de Psicologia da Saúde" />
<meta name="citation_title" content="Adaptação e estudo da escala de valor da saúde" />
```

Collaboration among HAL
and Google Scholar - and role
of repository administrators

HAL & Google Scholar teams have long collaborated

- HAL and Google Scholar teams have been working closely for many years
- We collaborate closely on indexing related features & infrastructure
- This collaboration has been very successful in ensuring broad coverage of French repositories, increasing global visibility for French research outputs
- We continue to work together to ensure repositories are indexed comprehensively and accurately as scale of repository content grows exponentially

Role of repository administrators

- Accurate & effective indexing requires both platform and repository level effort
- In the rest of the slides, we present what repository administrators can do to make sure hosted articles are indexed well
 - Focus on automated, sustainable approaches rather than manual fixes, which are increasingly difficult at the scale of most current institutional repositories

The crucial role of
accurate metadata for
indexing

Inaccurate metatags prevent indexing

- Indexing in Scholar requires accurate metatags
 - e.g. the order of author metatags in the HTML source controls the author order that appears in Scholar search results.
- Incorrect metadata results in unhappy authors & items not ranked as they should be due to missing citations.
- The indexing system automatically detects sites with frequent metadata errors and stops including them.

Metatags need to match the version of record:

PSICOLOGIA, SAÚDE & DOENÇAS 2009, 10 (2), 217-225

ADAPTAÇÃO E ESTUDO DA ESCALA DE VALOR DA SAÚDE

Filipa Pimenta¹, Isabel Leal^{1,2} & João Maroco^{1,2}

¹Unidade de Investigação em Psicologia e Saúde, I&D

² ISPA

RESUMO: O presente estudo pretende averiguar se pessoas com diferentes comportamentos de consumo de tabaco valorizam a sua saúde de forma discrepante; é igualmente objectivo desta investigação adaptar a Escala de Valor da Saúde (Lau, Hartman, & Ware, 1986) à população Portuguesa. Para tal aplicou-se o instrumento a uma amostra de 380 estudantes universitários (fumadores regulares, ocasionais e ex-fumadores). Os resultados demonstram que pessoas com diferentes padrões de consumo tabágico atribuem um valor elevado à sua saúde, não existindo diferenças significativas entre os três grupos ($F=1,594$; $p=0,205$). O estudo das características psicométricas do

→ [view-source:http://repositorio.ispa.pt/handle/10400.12/1089](http://repositorio.ispa.pt/handle/10400.12/1089)

```
<meta name="citation_author" content="Pimenta, Filipa" />
<meta name="citation_author" content="Leal, Isabel Pereira" />
<meta name="citation_author" content="Maroco, João" />
<meta name="citation_issn" content="1645-0086" />
<meta name="citation_journal_title" content="Psicologia, Saúde & Doenças" />
<meta name="citation_date" content="2009" />
<meta name="citation_abstract_html_url" content="https://repositorio.ispa.pt/handle/10400.12/1089" />
<meta name="citation_lastpage" content="225" />
<meta name="citation_language" content="por" />
<meta name="citation_pdf_url" content="https://repositorio.ispa.pt/bitstream/10400.12/1089/1/PSD%202009%2010%282%29%20217-225.pdf" />
<meta name="citation_keywords" content="Valor; Saúde; Fumadores; Ex-fumadores; Escala; Value; Health; Smokers; Ex-smokers; Scale; article" />
<meta name="citation_publisher" content="Sociedade Portuguesa de Psicologia da Saúde" />
<meta name="citation_title" content="Adaptação e estudo da escala de valor da saúde" />
```

Common metatag errors for repository publications

- Incorrect date listed in publication date metatags, most commonly listing the online/upload date as the publication date
- Missing authors in citation_author metatags, most commonly only including authors from institution or only listing the first author of a publication
- Incorrect author order, most commonly adding author(s) from the institution first

How to ensure best
possible metadata on the
HAL platform

HAL metadata best practices begins with deposit

- Metadata automation at deposit helps avoid common metadata errors
- Add DOI at upload whenever available
 - Adding DOI initially retrieves metadata from Crossref for improved accuracy
- When DOI doesn't exist, metadata can be extracted from uploaded PDF directly

Submit the file(s) Display detailed view

submit a file or/and complete your submission from an external identifier.

Transfer the files

Select the files...

Load metadata from a document identifier

With the document identifiers we are able to automatically get information on the document. The DOI, for example, is a code which appears on the editor's page of an article.



✓ The type of your document is 'Journal articles'.

In case of error, you can choose the document type in the section "Complete the document metadata."

✓ Nous avons complété votre dépôt à partir de l'identifiant :
10.1080/15362426.2017.1384769

Fulltext language, Title, Record date, Journal, Volume, Page number, Number, Identifiers



(s) ✓

document's metadata Display the complete

Document type * * Required fields

Title * You can add the title in multiple languages by clicking on +.

Metadata: what HAL repository administrators can do

- Reviewing metadata for possible errors, including any occasional errors caused by automated metadata extraction from PDFs
- Ensuring accurate metadata for items without DOIs, especially where library is hosting the version of record, e.g. theses and dissertations

How to ensure best
possible metadata on the
DSpace platform

Fixing common DSpace metadata issues

- Several common errors can often be fixed systematically via code patches or platform upgrades
 - Online/upload date listed as publication date in metatags
 - Incorrect author order in author metatags
- Other common metadata errors currently require manual corrections, e.g. missing authors in author metatags

Incorrect publication dates in citation_date metatag

- Test by comparing date citation_date tag with date in the version of record. If date listed in citation_date tag is later than the version of record, likely online date is being used as publication date.
- Systematic fix: patch for DSpace repositories available at <https://github.com/DSpace/DSpace/pull/2294.patch>
 - More information here: <https://jira.duraspace.org/browse/DS-4104>

Incorrect author order in citation_author metatags

- Systematic fix: If using DSpace versions 5.0, 5.1, 5.2, and 5.3, patch for DSpace repositories available at <https://github.com/DSpace/DSpace/pull/999>
 - More information here: <https://jira.duraspace.org/browse/DS-2679>
 - Upgrading to DSpace v5.4 or above will also fix this problem
- Manual fix: Adjust author metatag order to match author order in the version of record

Checking your repository's coverage

How to do a Scholar coverage check for your repository:

What **doesn't** work

- The result count of searching your repository site (“site:XXX”) is not an accurate indicator of Scholar coverage
- This is because the number listed in Scholar search results for a site only applies to the primary links – and as described earlier, the repository content is often not the primary link (in “All XX versions”)

Google Scholar

site:dspace.myuniversity.org



Not effective!

How to do a coverage check for your repository: what **does** work

- Search in Scholar for titles of several dozen **randomly** selected items across the repository and see if these papers are included
- Be sure to check the “All XXX versions” link as well, as often the repository version will not be the primary link

Google Scholar "3D sound scattering by rigid barriers in the vicinity of tall buildings"

Articles

Any time
Since 2019
Since 2018
Since 2015
Custom range...

Sort by relevance
Sort by date

[3D sound scattering by rigid barriers in the vicinity of tall buildings](#)
[L Godinho, J António, A Tadeu](#) - Applied Acoustics, 2001 - Elsevier
The boundary element method (BEM) is used to evaluate the acoustic scattering of a three-dimensional (3D) sound source by an infinitely long rigid barrier in the vicinity of tall buildings. The barrier is assumed to be non-absorbing and the buildings are modeled as an infinite barrier. The calculations are performed in the frequency domain and time signatures are obtained by means of inverse Fourier transforms. The 3D solution is obtained by means of Fourier transform in the direction in which the geometry does not vary. This requires ...
☆ 📄 Cited by 53 Related articles **All 7 versions**

[3D sound scattering by rigid barriers in the vicinity of tall buildings](#)

[L Godinho, J António, A Tadeu](#) - Applied Acoustics, 2001 - Elsevier

The boundary element method (BEM) is used to evaluate the acoustic scattering of a three-dimensional (3D) sound source by an infinitely long rigid barrier in the vicinity of tall buildings. The barrier is assumed to be non-absorbing and the buildings are modeled as an ...

☆ 📄 Cited by 53 Related articles

[3D sound scattering by rigid barriers in the vicinity of tall buildings](#)

L Godinho, J Antonio, A Tadeu - Applied Acoustics, 2001 - eilibrary.ru

The boundary element method (BEM) is used to evaluate the acoustic scattering of a three-dimensional (3D) sound source by an infinitely long rigid barrier in the vicinity of tall buildings. The barrier is assumed to be non-absorbing and the buildings are modeled as an ...

📄

[3D sound scattering by rigid barriers in the vicinity of tall buildings](#)

L Godinho, J António, A Tadeu - 2001 - estudogeral.sib.uc.pt

The boundary element method (BEM) is used to evaluate the acoustic scattering of a three-dimensional (3D) sound source by an infinitely long rigid barrier in the vicinity of tall buildings. The barrier is assumed to be non-absorbing and the buildings are modeled as an ...

📄

Additional Scholar indexing guidelines and resources

Google Scholar guidelines and resources for repositories

1. Google Scholar inclusion guidelines & troubleshooting guidelines

<https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>

<https://scholar.google.com/intl/en/scholar/inclusion.html#troubleshooting>

2. “Indexing Repositories: Pitfalls & Best Practices” presentation from 2015 Open Repositories conference

<https://www.or2015.net/wp-content/uploads/2015/06/or-2015-anurag-google-scholar.pdf>

Final thoughts

- Repositories are a crucial pathway for scholarly access
 - But what can't be discovered can't be built upon
- Platform automation for accurate metadata is key for sustainable, scalable repositories
- We need to work together to help researchers everywhere advance from a shared frontier of knowledge

Thank you for joining us!
Questions?

mwestin@google.com