# Expanded citations and projections of concepts*

Marcin Skulimowski
Faculty of Physics and Applied Informatics
University of Lodz
Pomorska 149/153
90-236 Lodz, Poland
mskulim@uni.lodz.pl

## ABSTRACT

In our recent paper, we proposed a new kind of citations, called the *expanded citations*, which link scientific papers and concepts from them. The expanded citations are represented in RDF and can be processed by machines. In this paper, we use the expanded citations to introduce *projections of concepts* which can be useful in searching for publications. The analysis of the projections and their time evolution gives a knowledge about the role and the significance of the concept in a given domain.

## Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods

## Keywords

citation relation, semantic publishing, digital libraries

## 1. INTRODUCTION

Bibliographic citations are very important in scientific activity. They can be used to estimate the quality of the research ([1],[2],[12]). The problem with citations, however, lies in the fact that they usually are treated equally, while, in fact, there are many reasons why authors cite publications of others ([5],[9],[11]). A growing body of literature has tried to describe precisely citations and, in consequence, the relationship between publications ([4],[10]). To this end, terms from special ontologies can be used (see e.g. [6]). Accordingly, citation networks can be represented in RDF (Resource Description Framework) and become machine-understandable. However, to make this possible, we have to determine functions of citations in publications. This can be done automatically to some extent [3]. Above all, the function of a citation can be determined by the author of a publication. In fact, the author can do more. Namely, he can *expand* some of the citations. Note that a citation is often related to the usage of some concept (idea, formula, method) from a cited article in a citing article. And this concept can be "included" in the citation. In this way we come to the so-called *expanded citations* which we proposed in our recent paper [8].

In this paper, we propose an application of expanded citations. Namely, we show that expanded citations allow for more precise search of scientific publications. Today, when we want to find articles related to some topic, we usually use a search engine. As a search result we obtain a list of articles in which some word, related to the issue, was found. In order to decide whether an article is interesting for us or not we have to read or at least browse it. Expanded citations will enable search for a well-defined concept (identified by a URI - Uniform Resource Identifier). Moreover, as a search result we will obtain a collection of articles where the relationship between each article and the concept is defined by a term from some ontology. It will be possible to narrow down the results to a set of articles that are only in a certain relation to the concept. We call this set a *projection of the concept*. For example, we may be interested only in articles in which some concept was *generalized* or *criticized*. This short paper is organized as follows. In Section 2 the notion of expanded citations is presented. Projections of concepts are introduced and briefly discussed in Section 3. The paper ends with discussion and the future directions of the present work.

## 2. EXPANDED CITATIONS

Throughout this paper, we use a term *concept* to refer to any entity (part) of a scientific article named with a URI. Such a URI can be obtained using a URL of the article (see [8]). Thus, each *concept* is a *resource* according to RDF terminology. Let us now consider two articles A and B and a standard citation $A \rightarrow B$ ($A$ cites $B$). In our recent paper, we introduced the following definition [8]:

*Definition 1.* We say that a citation $A \rightarrow B$ is *expandable* if there exist concepts $C_A$ (from $A$) and $C_B$ (from $B$), relations $r$, $r_A$, $r_B$ represented by object properties from some ontology $O$ and the following RDF triples (links)[1]:

$$C_A \, r \, C_B. \tag{1}$$

$$A \, r_A \, C_A. \tag{2}$$

$$B \, r_B \, C_B. \tag{3}$$

*Dedicated to the memory of my father, who recently passed away.



[1]Throughout this paper, we use Notion 3 syntax for RDF (`http://en.wikipedia.org/wiki/Notation3`). In this definition, for simplicity, we use letters for resources.

The set of triples (1-3) we will call an *expanded citation* (see Fig. 1).

Standard citations are included in *citation networks* [1]. Expanded citations are included in the so-called *concept networks* [7].

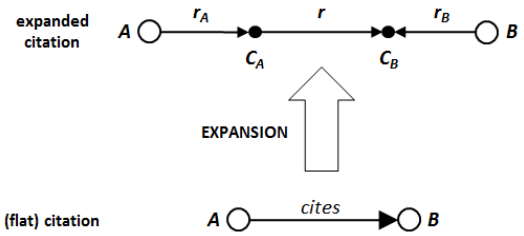


**Figure 1: Expanded citation [8]**

### *Example 1*

Let us consider *Kijowski's distribution for the time of arrival* (shortly *Kdistr*). This concept was introduced in the following paper:

Kijowski, J., *On the time operator in quantum mechanics and the Heisenberg Uncertainty relation for energy and time*, Reports on Mathematical Physics 6 (3), p. 361-386.

The paper has the following URL:

`http://dx.doi.org/10.1016/S0034-4877(74)80004-2`

The concept *Kdistr* is represented in this paper by the formula with number 9. So we can assign to it the following URI:

`http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9`

We take into account the following set of articles related to the concept *Kdistr*:

```
1-http://dx.doi.org/10.1006/spmi.1997.0544
2-http://dx.doi.org/10.1103/PhysRevA.58.4336
3-http://link.aps.org/doi/10.1103/PhysRevA.61.012104
4-http://pra.aps.org/abstract/PRA/v61/i2/e022118
5-http://dx.doi.org/10.1016/S0370-1573(00)00047-8
6-http://arxiv.org/abs/quant-ph/0102005
7-http://m.iopscience.iop.org/1751-8121/43/50/505303
```

Each of these articles contains a standard citation referring to the article (dentoted by 0) in which the concept *Kdistr* was introduced. All these citations refer to the concept *Kdistr*. They can be expanded using terms from SACO, the Scientific Article Content Ontology.[2] This draft ontology contains a set of object properties enabling characterization of what is done within a scientific publication. For example, we can say that an article *analyzes*, *generalizes* or *uses* some concept. The expanded citations obtained for *Kdistr* can be found in Appendix A.

## 3. PROJECTIONS OF CONCEPTS

As it follows from the previous section, an expanded citation consists of three RDF triples (see Definition 1). In this section we limit ourselves to the following triples:

$$P\ r\ C. \tag{4}$$

[2]http://purl.org/lyr/saco

where $P$ is a publication, $C$ is a concept from $P$ and $r$ is a named relation (represented by an *object property*) between $P$ and $C$. The same concept $C$ can appear in many articles. It may be useful to consider all publications referring to this concept. We introduce the following definition.

*Definition 2.* Let $C$ be a concept and $O$ be an OWL ontology. The set $P(C)$ of all publications referring to $C$ is defined as follows:

$$P(C) := \{P \mid \exists r \in O\ .\ P\ r\ C\}. \tag{5}$$

Notice that $P(C)$ is not a set of publications in which some word (e.g., a name related to the concept $C$) can be found using full-text search. $P(C)$ contains publications for which relations to the concept $C$ are precisely defined by $r$. Each such a relation $r$ corresponds to some subset of $P(C)$.

*Definition 3.* A *projection of a concept* $C$ on an object property $r \in O$ is a set $r$-$C$ defined as follows:

$$r\text{-}C := \{P \in P(C) \mid P\ r\ C\}. \tag{6}$$

As we can see from the definition, $r$-$C$ contains all publications which are linked with $C$ by the relation $r$. Note that $r$-$C$ can be defined formally in OWL as a restriction containing an *owl:hasValue* constraint. It describes a class of all individuals (in our case — publications) for which the property $r$ has at least one value semantically equal to the concept $C$. From the last definition it follows that the family of sets $\{r\text{-}C : r \in O\}$ is a *cover* of $P(C)$, i.e:

$$P(C) = \bigcup_r (r\text{-}C) \tag{7}$$

Using properties from the SACO ontology we can obtain

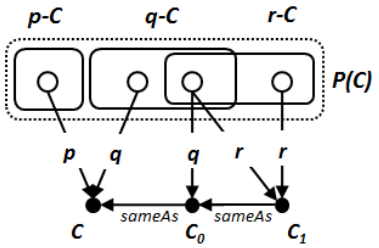


**Figure 2: Projections of $C$ as a cover of $P(C)$.**

various projections of a concept $C$ e.g.:

*introduces-C* - contains a publication (or publications) in which $C$ was introduced for the first time.
*uses-C* - contains publications in which $C$ is used.
*generalizes-C* - contains publications in which $C$ is generalized.
*givesExampleOf-C* - contains publications in which examples of $C$ are given.

### *Example 2*

Using the expanded citations obtained for the Kijowski's distribution (see Appendix 1) we can obtain the following sets[3]:

[3]Note that, we do not consider projections of *Kdistr* (which contain all publications) but their subsets.

$\{0\} = introduces\text{-}Kdistr$
$\{1, 2, 5\} \subset analyzes\text{-}Kdistr$
$\{4, 6, 7\} \subset generalizes\text{-}Kdistr$
$\{3, 7\} \subset obtainsDifferently\text{-}Kdistr$

We can see from the above examples that projection spaces may intersect. It is obvious taking into account that the same concept can be "treated" in different ways in one publication. For example, our concept *Kdistr* is *generalized* and *obtained differently* in publication 7. In consequence, the family of sets $\{r\text{-}C : r \in O\}$ is a *cover* of $P(C)$ but not its *partition*. In some cases, one projection may be a subset of another one. For example, it seems to be obvious that if a publication *introduces* some concept $C$ then it also *analyzes* this concept. Consequently: $introduces\text{-}C \subset analyzes\text{-}C$. In general, we can formulate the following lemmas.

LEMMA 1. *Let $r_1$ and $r_2$ be two object properties defined in some OWL ontology. Then:*

$$r_1\ rdfs:subPropertyOf\ r_2 \Rightarrow r_1\text{-}C \subset r_2\text{-}C$$

*where $C$ is a concept.*

PROOF. Let us consider $P \in r_1\text{-}C$. From the definition of $r_1\text{-}C$ it follows that there exists an RDF triple $P\ r_1\ C$. Because $r_1$ is a subproperty of $r_2$ there exists an RDF triple $P\ r_2\ C$. This means that $P \in r_2\text{-}C$. □

Obviously the reverse implication is not true. One can imagine that for some concept $C$ and two properties $r_1$ and $r_2$ we have $r_1\text{-}C \subset r_2\text{-}C$. However, it does not necessarily follow from this inclusion that $r_1$ is an *rdfs:subproperty* of $r_2$.

LEMMA 2. *Let $r_1$ and $r_2$ be two object properties defined in some OWL ontology. Then:*

$$r_1\ owl:propertyDisjointWith\ r_2 \Rightarrow r_1\text{-}C \cap r_2\text{-}C = \emptyset$$

*where $C$ is a concept.*

PROOF. Assume to the contrary that there is a publication $P$ such that $P \in r_1\text{-}C \cap r_2\text{-}C$. From the definition of a projection it follows that the statements $P\ r_1\ C$ and $P\ r_2\ C$ are true. But this contradicts our assumption that $r_1$ is disjoint with $r_2$. □

## 3.1 Evolution of projections

Projections of concepts will change over time as new publications will appear. The analysis of both the projections $p{-}C$ and their time evolution can provide information about the role and importance of $C$ in a given domain. Important concepts will have a number of non-empty projections. Besides, these projections will contain more and more publications over time. In order to illustrate the time evolution of projections we propose special diagrams called *projection diagrams* (see Fig. 3). A projection diagram represents time evolution of several projections. The horizontal lines correspond to projections. The points on these lines represent publications. What is very important, the projection diagram includes a time axis. The publications are arranged chronologically, from the oldest (on the left) to the youngest

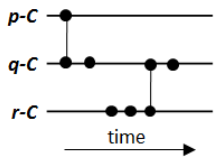


**Figure 3: Projection diagram for some concept $C$.**

(on the right). In the case when some publication belongs to two (or more) projections the points representing this publication are connected by a vertical line. Empty projections are not taken into account.

### *Example 2*

Let us consider the concept *Kdistr* and the papers from Example 1. Taking into account their publication dates we can obtain the *projection diagram* for Kijowski's distribution. It is presented below in Figure 4.

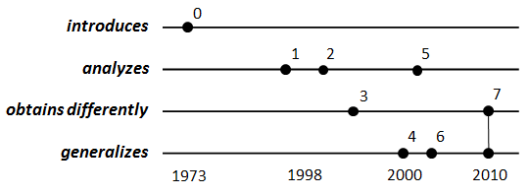


**Figure 4: Projection diagram for Kijowski's distribution.**

Let us examine the diagram in more detail. The considered concept *Kdistr* was introduced in paper 0. There are three generalizations of the concept (papers 4, 6, 7). Moreover, in papers 3 and 7 the concept was obtained in a different way. In three papers (1, 2, 5) the concept was analyzed. There is also one paper (7) in which the concept was generalized and obtained differently. When we take into account the timeline of the concept we can observe that it was introduced in 1973 and the generalizations of it were obtained about 30 years later. It suggests that the concept is very important in the domain.

## 3.2 Discussion and future work

Projections of concepts allow to decompose the set $P(C)$ of all publications related inaccurately to a concept $C$ into various sets for which the relationship is precisely defined. This decomposition will make it easy to find publications related to the concept $C$ in some clearly defined way. Using the union and the intersection of projections we are able to define very accurately the set of publications we are interested in. Finding publications contained in this set will be more efficient because the projections of concepts can be generated automatically using machine-understandable expanded citations.

To further our research at first we plan to develop and harmonize the vocabulary used in expanded citations. Obviously, not every standard citation can be expanded. We intend to investigate how large is the class of citations for which the expansions exist. The current study was limited to publications from one domain (quantum mechanics). Future work will explore expanded citations and projections

of concepts in various domains. We are currently developing a web tool supporting the creation of expanded citations and search for publications with the help of projections of concepts.

# 4. REFERENCES


[1] L. Egghe and R. Rousseau. *Introduction to Informetrics: quantitative methods in library, documentation and information science*. Elsevier Science Publishers, 1990.

[2] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United Stated of America (PNAS)*, 102(46), 2005.

[3] A. D. Iorio, A. Nuzzolese, and S. Peroni. Towards the automatic identification of the nature of citations. In *Proceedings of 3rd Workshop on Semantic Publishing (SePublica 2013), CEUR Workshop Proceedings 994: 63-74*, 2013.

[4] B. Jörg. Towards the nature of citations. In *Poster Proceedings of the 5th International Conference on Formal Ontology in Information Systems*, 2008.

[5] M. J. Moravcsik and P. Murugesan. Some results on the function and quality of citations. *Social Studies of Science*, 5(1):86–92, 1975.

[6] S. Peroni and D. Shotton. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Web Semantics*, 17:33–43, 2012.

[7] M. Skulimowski. From linked data to concept networks. In *Theory and Practice of Digital Libraries 2013, Communications in Computer and Information Science, Vol. 416*, pages 77–88, 2014.

[8] M. Skulimowski. On expanded citations. In *Proceedings of the i-KNOW*, 2014.

[9] I. Spiegel-Rusing. Bibliometric and content analysis. *Social Studies of Science*, 7(1):97–113, 1977.

[10] S. Teufel, A. Siddarthan, and D. Tidhar. An annotation scheme for citation function. In *Proceedings of 7th SIGdial Workshop on Discourse and Dialogue*, pages 80–87, 2006.

[11] M. Weinstock. Citations indexes. *Encyclopedia of Library and Information Science*, 5:16–40, 1971.

[12] H. D. White. Citation analysis and discourse analysis revisited. *Applied Linguistics*, 25:89–116, 2004.


# APPENDIX

Below you can find the expanded citations obtained for the Kijowski's distribution and the papers from the Example 1. For simplicity we omit the following triple (which is included in each of the citations):

```
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2>
sac:introduces
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 1:

```
<http://dx.doi.org/10.1006/spmi.1997.0544>
sac:analyzes
<http://dx.doi.org/10.1006/spmi.1997.0544#10>.
<http://dx.doi.org/10.1006/spmi.1997.0544#10>
owl:sameAs
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 2:

```
<http://dx.doi.org/10.1103/PhysRevA.58.4336>
sac:analyzes
<http://dx.doi.org/10.1103/PhysRevA.58.4336#10>.
<http://dx.doi.org/10.1103/PhysRevA.58.4336#10>
owl:sameAs
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 3:

```
<http://link.aps.org/doi/10.1103/PhysRevA.61.012104>
sac:obtainsDifferently
<http://link.aps.org/doi/10.1103/PhysRevA.61.012104
#Pi_psi_t>.
<http://link.aps.org/doi/10.1103/PhysRevA.61.012104
#Pi_psi_t>
owl:sameAs
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 4:

```
<http://pra.aps.org/abstract/PRA/v61/i2/e022118>
sac:introduces
<http://pra.aps.org/abstract/PRA/v61/i2/e022118#16>.
<http://pra.aps.org/abstract/PRA/v61/i2/e022118#16>
sac:generalizes
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 5:

```
<http://dx.doi.org/10.1016/S0370-1573(00)00047-8>
sac:analyzes
<http://dx.doi.org/10.1016/S0370-1573(00)00047-8#2>.
<http://dx.doi.org/10.1016/S0370-1573(00)00047-8#2>
owl:sameAs
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 6:

```
<http://arxiv.org/abs/quant-ph/0102005>
sac:introduces
<http://arxiv.org/abs/quant-ph/0102005#19>.
<http://arxiv.org/abs/quant-ph/0102005#19>
sac:generalizes
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```

Expanded citations for paper 7:

```
<http://m.iopscience.iop.org/1751-8121/43/50/505303>
sac:introduces
<http://m.iopscience.iop.org/1751-8121/43/50/505303
#65>.
<http://m.iopscience.iop.org/1751-8121/43/50/505303
#65>
sac:generalizes
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
<http://m.iopscience.iop.org/1751-8121/43/50/505303>
sac:obtainsDifferently
<http://m.iopscience.iop.org/1751-8121/43/50/505303
#62>.
<http://m.iopscience.iop.org/1751-8121/43/50/505303
#62>
owl:sameAs
<http://dx.doi.org/10.1016/S0034-4877(74)80004-2#9>.
```