

Rick Szostak
University of Alberta, Edmonton, Alberta, Canada

Poly-coordination (CAIS Paper)

Abstract: A structured approach to synthetic classification is feasible and allows us to achieve the best of both pre-coordination and post-coordination. In particular it allows both precision and ease of use. Poly-coordination is particularly well-suited to visualization technologies. It links classification research with information retrieval.

Résumé: Une approche structurée de classification synthétique est réalisable et nous permet d'obtenir le meilleur de deux approches, soit de pré-coordination et de post-coordination. En particulier, elle permet à la fois la précision et la facilité d'utilisation. La poly-coordination est particulièrement bien adapté aux technologies de visualisation. Elle relie la recherche de classification avec la récupération d'information.

Introduction

This paper makes four related points:

- It is possible to pursue an approach to classification that combines the best of pre-coordination and post-coordination. I call this approach poly-coordination. The key here is to make the order of search terms matter.
- We can then have the advantages attributed in the literature to both pre- and post-coordination.
- Poly-coordination may be an avenue that can link the distinct fields of classification research and information retrieval
- Poly-coordination is also well-suited to the employment of new visualization technologies.

Poly-Coordination is Feasible

The classification literature distinguishes pre-coordinated classifications (from which classifiers must choose from among a predetermined set of subject headings) from post-coordinated (in which classifiers can synthesize simpler terms in a classification). The literature assumes that we must choose between one or the other approach, and describes the strengths and weaknesses associated with each (Svenonius 1993, Jacob 2004, Sauperl 2009). But is it possible to combine the best of both (Szostak 2015)?

The literature attributes a lack of precision to post-coordination. In particular it is often noted that a post-coordinated approach fails to distinguish "philosophy of history" from "history of philosophy." Yet this is – I would argue – not an inherent shortcoming of post-coordinated systems in general but rather an artifact of the way these have been approached and utilized.

I worked with a group of computer science students in 2015 to apply structured search algorithms to the Basic Concepts Classification (Szostak 2013). They developed a set of algorithms that could place different weights on the order of search terms when presenting results. It is thus technically quite straightforward to privilege the order of terms in a search interface. A user wanting “History of philosophy” can then input the search terms (history)(of)(philosophy) in that order, and be rewarded with precisely what they seek. If,

however, the user was doing an exploratory search for any interaction between history and philosophy they could opt for a different search algorithm that would give them both history of philosophy and philosophy of history.

In sum, the oft-stated conjecture that a synthetic approach necessarily lacks precision can now be shown to be empirically untenable. A poly-coordinated system employs synthesis but in a structured fashion such that the order of search terms matters. Poly-coordination can be justified theoretically with reference to the literature cited above on pre- and post-coordination: once the conjecture regarding synthesis and precision is overturned the result follows that it is possible to combine the best attributes of pre- and post-coordination. Szostak (2013) surveys other theoretical justifications – including the important insight that such an approach allows facet analysis without requiring facet indicators.

Poly-coordination is Superior to Pre-Coordination and Post-Coordination

We can thus have the advantages of post-coordination while also achieving the precision associated in the literature with pre-coordination. And this is a powerful combination for the user need not master a complex pre-coordinated system in order to achieve precise results. The user should then be able to find desired documents more quickly than with either strict pre- or post-coordination, for they can input combinations of simple search terms that are derived directly from their search query (rather than having to search for a pre-coordinated heading) and yet be rewarded with precise results.

The key to poly-coordination is that both classifier and user employ sentence-like structures in generating synthetic strings. A book about dogs biting mail carriers will be treated as (dogs)(bites)(mail carrier) by both.

Poly-coordination and Information Retrieval

There is an obvious connection here to the information retrieval literature. Poly-coordination is feasible only in concert with appropriate search algorithms (see above). The field of information retrieval is increasingly skeptical of the "bag of words" assumption: that it makes sense to search for search terms independently. Scholars of information retrieval are thus looking at ways to structure search queries so that search terms are combined (Wallach 2006). Now of course information retrieval has long eschewed controlled vocabulary in favor of keyword searching. But while this has proven useful for many tasks it does not provide the necessary precision for the kinds of detailed searches that academics among others often wish to make. As information retrieval experts try to figure out how best to structure multi-term searches, the time is ripe for classificationists to suggest some recourse to controlled vocabulary: At the least keyword searches might be employed to guide users to relevant places within a classification system. Cleverly and Burnett (2015) note that it is quite feasible to combine automatic and manual search techniques. But retrieval experts will only be interested in an extremely user-friendly form of controlled vocabulary. This a poly-coordinated approach provides for it allows the user to structure a search from sentence-like combinations of basic (everyday) concepts: like (dog)(bites)(mail carrier). And a digital thesaurus could be employed to translate if the user employed the wrong search term (we just don't call them mailmen anymore).

We thus at the same time that we transcend the false dichotomy between post- and pre-coordination can potentially transcend an exaggerated dichotomy between subject and keyword searching. The key here is to recognize that subject searching need not be so very very hard. We have made subject searching difficult by forcing the user to navigate ad hoc and complex pre-

coordinated classification systems. We can instead give them a straightforward poly-coordinated system that they can master almost instantaneously (if it is wedded to an exhaustive thesaurus, which it certainly could be – as was noted by conference participants manual and automatic techniques can be harnessed to this task as well).

Poly-coordination and Visualization

An exciting area of classification (and information retrieval) research involves visualization techniques. Visual interfaces can visually suggest new avenues for search to the user not satisfied with their first hits (Gnoli and Cheti 2013). Visualization can be useful even with pre- or post-coordinated classifications, or with bag-of-words searching. But visualization can do even greater wonders in concert with a poly-coordinated system, for then a variety of alternative combinations can be suggested: A user seeking information regarding dogs biting mail carriers can be advised of a variety of related possibilities: Dogs kissing mail carriers, dogs biting florists, cats biting mail carriers, mail carriers biting dogs. These can be visually linked to the original search string, with different fonts or colors employed to capture different types of change to the original search string: to the subject or object or verb or the order of terms.

The key here is that a poly-coordinated system instantiates a “web of relations” among things and the influences these exert on each other. Pre-coordinated systems provide combinations of some sets of things and (sometimes) relationships – but since individual subject headings are themselves complex combinations of search terms, it is much harder for a computer interface to identify other subject headings that are related along diverse dimensions. Post-coordinated systems allow terms to be combined but without specifying relationships among these. Poly-coordination, by identifying subjects, objects, and verbs (adverbs and adjectives are also possible) – and thus mimicking the way that humans actually voice questions – places any search query within a wider array of influences. As in the examples above, it is then straightforward for the visual interface to suggest diverse slight alterations to the original query. (This is possible, we might note, because the hierarchies of things and adverbs/adjectives can be flat and logical, and the combinatorial structure of verb-like terminology straightforward in a poly-coordinated scheme [Szostak 2013]; visual interfaces will struggle with classifications that have gaps in hierarchies, or do not follow straightforward rules in establishing subclasses – such as when recycling is treated as a subset of garbage because there is no other place for this action term.)

Likewise a user curious about how A affects B can be introduced to a longer causal chain: influences on A, effects of B. This sort of connection is one of the key foci of the literatures on literature-based discovery or serendipity: many key advances in the history of science have resulted from combining understanding in one field of how A affects B with understanding in another field of how B affects C in order to identify an important but unrecognized connection between A and C. The user regularly confronted with visual representations of such connections is much more likely to produce important discoveries (and not just in science but in technology or art or indeed any sort of creative process).

And poly-coordination coupled with visualization techniques potentially has an important educational role (and may thus be of particular interest to school librarians). It is hard to acquaint students at any level with the LCC or DDC. But a poly-coordinated system is easy to teach. It involves short and logically organized hierarchies of things, verb-like relationships and adverbs and adjectives. These, as noted above, are easily visualized. And queries are structured in a sentence-like format that is easily appreciated. And the approach has the added benefit of

actually representing how human understandings are structured: as a finite set of relationships among a finite set of things. We could thus potentially teach grade school students the structure of knowledge *and* how to search this structure. And a visualization interface that allowed them to pursue their curiosity in unimagined directions would actually be fun (Wow; I never thought about mail carriers biting dogs).

Further Research

Though the feasibility of poly-coordination and structured search has been established, there is much scope for empirical evaluation of the efficacy of this approach.

References

- Cleverly, Paul & Simon Burnett (2015) The best of both worlds: Highlighting the synergies of combining manual and automatic Knowledge Organization methods to improve information search and discovery. *Knowledge Organization* 42:6, 428-44.
- Gnoli, Claudio, & Alberto Cheti (2013). Sorting documents by base theme with synthetic classification: the double query method. In A. Slavic, A.A. Salah, S. Davies (Eds.), *Classification and Visualization: Interfaces to Knowledge* (pp. 225-232). Würzburg: Ergon Verlag.
- Jacob, Elin K. (2004) Classification and Categorization: A Difference that Makes a Difference. *Library Trends* 52:3, pp. 515-40.
- Sauperl, Alenka (2009) Precoordination or not?: A new view of the old question. *Journal of Documentation* 65:5, pp. 817 – 833.
- Svenonius, Elaine (1993) Pre-coordination or not? Subject indexing in the 90s: principles and practices, paper presented at IFLA Satellite Meeting, Lisbon, 17-18 August.
- Szostak, Rick (2013) The Basic Concepts Classification.
<https://sites.google.com/a/ualberta.ca/rick-szostak/research/basic-concepts-classification-web-version-2013>
- Szostak, Rick (2015) “A Pluralistic Approach to the Philosophy of Classification” *Library Trends*,63:3, 591-614.
- Wallach, Hanna M. (2006). Topic modelling: Beyond ‘bag of words.’ In *Proceedings of 23rd International Conference on Machine Learning*. Pittsburgh.