Searching bibliographies

Alan Braslau
CEA, France

BibTeX identifies its database entries using unique keys and it is up to the creator of the database choose keys that an author can use to insert references using the `\cite` command. Many database handling tools such as `jabref` can autogenerate keys according to a particular scheme (such as ‘LastnameYear’ with an eventual suffix, a, b, c, …); Some website exports might use ‘JournalVolumePage’ or variants.

The challenge for the author is remembering such keys when writing and is complicated by the fact that autogenerated keys might get changed as the database is sorted differently or as new entries are added.

A more natural way of thinking involves different schemes for different writers. Some might remember author names, journal names, years, …, others might remember the thesis advisor's name (rather than the first author's name), still others might remember the title of a work. The database might contain `keywords` that can be used to identify works. This is why it is eminently useful to be able to search the database using some logical scheme.

Here is an example: (Darwin, 1859)

```
\cite[dataset::match(author:Darwin and year:1859 and
title:Origin)]
```

The syntax uses the function `match()` working on a logical expression.

> I have suggested that ‘`match`’ can be made optional, as `match()` is the only recognized function at present. Hans would like to keep the syntax open to extension as we will surely, eventually find the need for other schemes.

The system is quite flexible and can mix `match()` strings with classical BibTEX-style keys.

1 Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life*. London, John Murray

Pitfalls

One must be quite careful in the specification of the logical expression as an incomplete or improperly qualified search might yield unexpected results. As an example

```
\cite [dataset::match(author:Darwin and year:1859)]
```

might (and should) find several published works (in fact 10) and

```
\cite [dataset::match(author:Darwin and title:Origin)]
```

could identify many different editions (in fact 530, counting translations)! These may or may not be the intended results.

Another possible source of confusion is that the dataset can be filled using

```
\usebtxdataset [dataset] [source]
```

that can be located **anywhere** in the source file(s). The compilation runs over several passes and the datasets are saved from pass to pass. As the search is over the entire dataset, different dataset namespaces must be used if one seeks isolation, for example between chapters in a book.