

## CCM tools: Deduplication Hands-on worksheet

---

### Introduction & Accessing CCM tools

This worksheet focuses on the on-the-fly deduplication options within CCM tools. We would encourage you to work in pairs or small groups to discuss the results of the example searches.

The worksheet aims to help you:

- ❖ Gain a clearer understanding of the on-the-fly deduplication options available;
- ❖ Consider how you might apply them in practice;
- ❖ Gain a better appreciation of the variability of some of the data and to think about the impact this may have on your use of the tools.

As a reminder:

- ❖ To access the CCM tools interface go to: <http://ccm.copac.ac.uk>

Select the 'Sign in' option towards the top right of the screen. Login using your personal username and password.


#### Note:

We are trying to highlight some of the things you may need to be aware of as you decide whether to use the on-the-fly deduplication options and, if so, which option to choose for a particular collection analysis activity. That means some of the examples are for more complex cases, but they won't all be like this and many searches may give quite straightforward results!

## On-the-fly deduplication options

Within the CCM interface there are four on-the-fly deduplication options:

- ❖ ISBN deduplication
- ❖ Level 1: uses Date, Title, Pagination, Edition, Author, Publisher
- ❖ Level 2: uses Date, Title, Author, Publisher
- ❖ Level 3: uses Title, Author

Next to the 'Personalise result display' heading, selecting the  button will take you to the Help relating to the deduplication options.

Read through the help before you start searching to get a feel for what the deduplication options are doing. If you have any questions about these options please let us know.

## Testing the deduplication options

Use the Keyword search screen and try some of the example searches below, doing each chosen search five times:

1. Without using deduplication
2. Using each of the deduplication options in turn

For each search note the number of records in the result and look at the way in which the result set changes (or doesn't) as you use each form of deduplication.

As you do the searches and look at the results you might like to consider the following areas, the notes with each search suggest some questions to get you started:

- ❖ What are some of the pros and cons of each form of deduplication?
- ❖ In what situations might you want to use the different types of deduplication?

## Example Keyword searches.

The first couple of examples give a broad impression of the way the results can change using the deduplication options, the other two examples look in a bit more detail at the impact of the changes in the record displays. Feel free to try your own example searches to look at the effect of the deduplication in your area of interest.

1. **Author= camus, albert**  
A consistent pattern of result set size change with each different type of deduplication; perhaps what would be expected?
2. **Title= photosynthesis**  
The ISBN option gives a drop in result set size, which then goes up again with the level 1 match. Why might that be?  
The ISBN match will check all ISBNs, which could include set ISBNs; would that matter?

3. **Title: The American past : a survey of American history**  
ISBN match brings together records for separately catalogued volumes with single records representing multiple volumes. This doesn't happen with the title matches. Is this a problem? Or a benefit?  
If there is no ISBN in the records what do the results tell you about the likelihood of effective deduplication?  
The level 3 match brings together multiple editions (albeit not all). Why is that? Does this matter?
4. **Title= Victoria history of the county of Sussex Lewes**  
For this example look at the basic (unduplicated) search results, then compare them with the results of a search for:  
**Title= Victoria history of the county of Sussex**  
Quite a contrast in results – what might be an issue here?  
You might like to look at record 23 from the London Library.  
  
You could also try a search for:  
**Title= Victoria history of the counties of England.**  
**Library= Oxford**  
Look at record 3 from Oxford, this is a periodical record - what might be the issue here?

## If you have time...

On the attached sheet we'd appreciate your ideas as to the different contexts within which you might want to use the various deduplication options. This would help us to expand the information we provide and support colleagues in other institutions as they get to grips with this feature.



## Deduplication scenarios

We would appreciate any examples you might have of situations in which the different forms of on-the-fly deduplication might be valuable, with any pros and cons you might add to advise colleagues who were considering using these options.