You mean Keywords? 
How do I choose them?

Why Keywords? 
They seem to be useful :-) 

- "As any good library or information worker knows the accurate and consistent application of keywords can serve to enhance the content representation and retrieval of literature."

Why Keywords? 
Because, Words Matter!

- In fact, the way words are used is relevant, people working in the same areas share a common way of "saying things". For this reason:
  - "Topic detection techniques have been used to analyze the scientific literature to understand global research trends or see links and patterns amongst scientific documents."
  - "Example techniques include co-citation analysis, co-word analysis, co-author analysis, word frequency analysis, and recently probabilistic methods such as those based on latent Dirichlet allocation [2], or the use of self-organizing maps."

Now imagine… 
Do you feel the Stress?

- Close to 30 students…
  - each one of them wanting to work on a different topic
  - on which they are not experts (neither am I)
  - all of them anxious to have your help in figuring out:
    - are these the right keywords?
    - How do you manage so many different topics?
    - And how do you help students to "help" themselves?
Visualization helps…
[also a personal preference]

• A picture is worth a 1000 words, so, visualization should make sense.

I wish we could all be there!
Great location, beautiful Campus!

My path towards…
a possible solution!

• Started to see if bibliometric analysis could help me in some way…
• Mind that I would like to stay away from the topic of “bibliometric analysis”, but still, wanted to see if there was anything I could get from there.
• And I came across the R Package “Bibliometrix”

https://bibliometrix.org/

VOS vs Bibliometrix
The “vs” actually makes no sense ;-)
Actually, the way I use both, they complement each other!

• Bibliometrix is a wonderful tool for handling and processing massive amounts of data.
  • Bibliometrix is a R Package. This means the user may build on other R functions, Packages and scripting possibilities to enhance functionalities and automate frequent tasks.
  • It’s fast an easy to export whatever we like to VOSviewer.
• VOSviewer is a wonderful tool for visualization.
  • VOS imports bibliographic data (that we may export from R), and enables different types of analysis, involving for example, keywords (e.g.: co-occurrence) and references (e.g.: co-citation/ bibliographic coupling).
  • The graphics are great!

Then…
• While using Bibliometrix I started looking for other tools, and there are many tools out there!
• One was specially interesting:
  • VOSviewer - Visualization of Similarity

https://www.vosviewer.com/
So, what is the typical workflow?

**Data collection**

1. Define the Keywords to be used in the search.
2. Define and the search query (the logic combination of keywords to be used in the search)
3. Export to the search results using the BibTeX format (export the full bibliographic record)

**Making sense of the Data**

1. Import BibTeX file using Bibliometrix functions
2. Generate the summary of the imported data

So, what is the typical workflow?

**KEY**

( business AND entrepreneur* AND innovation )

**Video**

- See that you have “repeated” keywords...
- Using only author keywords (DE)
So, what is the typical workflow?

Making sense of the Data

3. Visualize Keywords (DE)

- Do they make sense to you?
- Is there any specific keyword that seems more valuable?

So, what is the typical workflow?

Making sense of the Data

Key: (business AND entrepreneur* AND innovation)

So, what is the typical workflow?

Making sense of the Data

4. Dig deeper on: “Business Model”

Extract from Scopus:

Key: (business and entrepreneur* and innovation AND “business model”) 30 docs

Import to R generate new graphic using VOS.

With few papers, if you want a picture of what you have, you need other tools!

Not much value on this visualization...

Remember, this is about exploring!

So, what is the typical workflow?

Making sense of the Data

4. UpSet: Visualizing Intersecting Sets (R Package)

https://caleydo.org/tools/upset/

We are seeing now intersecting keywords. This graphic is built from a sparse matrix where in each line has a “1” if the keyword occurs for that paper corresponding to that line.

Can we dig deeper?
So, what is the typical workflow?
Making sense of the Data

4. UpSet: Visualizing Intersecting Sets (R Package)

https://caleydo.org/tools/upset/

We are seeing now intersecting keywords. This graphic is built from a sparse matrix where in each line has a “1” if the keyword occurs for that paper corresponding to that line.

Can we dig deeper?

So, what is the typical workflow?
Steps...

1. Search SCOPUS/WOS
2. (Re-)Define (Keywords + Search Query)
3. Check Report Summary (Use biblioshiny to generate wordcloud)
4. Check Keyword Network + Paper
5. Iterate

UpSet: Visualizing Intersecting Sets + R

5. Dig deeper

We go back to R and look up the paper by filtering the column “DE” on the database with the keyword "social franchise"
Locate paper in table and get the actual document.
Is this the end? ;-)  

- Tipically, if I am supervising some student in his/her dissertation/thesis
  - I would like to provide the student with some relevant papers to read on the topic.
  - Here we have the tools to do it. We can use both:
    - co-citation
    - bibliographic coupling
  - If you use VOS it is quite easy, but not without some limitations…

---

Co-citation  
The contents of the exported file  

Exporting the papers to a “.csv” file allows the user to open it in excel…  
This seems useful, but there’s a catch!

---

Co-citation  
The contents of the exported file (another example)

The catch!

<table>
<thead>
<tr>
<th>VOS detected these papers as different, that is why they are both listed…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can you find the differences?</td>
</tr>
</tbody>
</table>

These two are the same papers

| --- |

---

Co-citation  
Using VOS for selecting co-cited papers  

Co-citation gives students the core papers for the area of the selected papers.

- This is likely the area where they want to do their research!
Co-citation
The contents of the exported file (yet another example)

The catch!

These two are the same papers

weiss, r., automatic food documentation and volume computation using digital imaging and electronic transmission (2010) j am diet assoc, 110, pp. 42-44

weiss, r., stumbo, p.j., divakaran, a., automatic food documentation and volume computation using digital imaging and electronic transmission (2010) j am diet assoc, 110, pp. 42-44

These two in red are the same papers, but both were listed!

mair, j., mart, i., social entrepreneurship research: a source of explanation, prediction, and delight (2006) journal of world business, 41 (1), pp. 36-44


Can you find the differences?

Issues to be aware of...
Keyword or Cited paper Similarity means (same sequences of chars)

• “Business model” seems to be the same as “Business Model “, but it is not!

• So, if you aim at:
  • just looking for the right keywords and just want to find nice papers to start your research you are ready to go!
  • doing some serious work and want to do a really serious co-citation analysis or keyword analysis...

  • The 1st step is to go through some uniformization effort. You may need some R scripting and even then, you will likely need several hours of manual work to fix the whole thing!

Bibliographic Coupling
Using VOS for selecting papers coupled by their references

• This is a nice approach if you have thousands of papers in your 1st search, which you think makes sense...

• So, you can use

  • co-citation to get older papers (the core of the research topic of your search)

  • bibliographic-coupling to get more recent papers

    • Mind that results may, sometimes, be disappointing... The reason is that these papers are not being selected for their number of citations... they are just “coupled” by their reference list!

Thank you!

Back to normal...

João José Pinto Ferreira | Associate Professor | Faculty of Engineering U.Porto | jjpf@fe.up.pt