**Forståelse af filers datastruktur**  
  
**For all sources**

* Each record should be parsed as a block.
* Within each block, extract the content from the lines starting with:
  + **Title:** Extract everything after this label until the next field label (for Embase/Medline) or until a new line (Web of Science).
  + **Abstract:** Same as above.

**Pubmed:**  
Each study in the pubmed file begins with a line that starts with a number followed by a period, such as "1."  
The title of the study follows a line labeled "Title".  
The abstract follows a line labeled "Abstract".  
Only these two fields should be extracted. All other lines, such as Authors, Journal, and Year, should be ignored.

**Embase:**  
Each study in the Embase file begins with a line like "Record [number]".  
The title is on the line labeled "Title:" and follows directly after the colon.  
The abstract is on the line labeled "Abstract:" and follows directly after the colon.  
Only the title and abstract should be extracted. All other metadata should be discarded.  
  
**Web of sciences:**   
Each study in the Web of Science file ends with a line containing only "ER".  
The title is indicated by the label "TI". The title may span multiple lines until the next label such as "SO", "AB", etc. These lines should be joined together as one title.  
The abstract is indicated by the label "AB". The abstract may also span multiple lines until the next label such as "SN", "PY", etc. These lines should be joined together as one abstract.  
Only the title and abstract fields should be retained. All other fields should be ignored.