

Conversion of Bibliography Entries

LaTeX is one of the most widely used typesetting programs. It is especially powerful in its capacity to express technical notation. For many books, including the textbook for this course, authors submit their work to the publisher in LaTeX format. LaTeX has conventions for many things such as equations, tables, and also, of interest here, bibliographical references.

If you have created HTML files, you know that doing so requires you to use a special set of instructions to get the desired formatting. LaTeX shares this characteristic, but is more powerful in that it lets you define and reuse your own special instruction sequences. It then sets things up and adjusts the details for your printed output. The price paid for this power is that LaTeX is not wysiwyg.

REFER is an earlier set of conventions developed specifically for bibliographies. In this assignment, you will be in the position of a person whose company or agency has a set of bibliographic entries in the REFER format that must be converted to LaTeX format.

Examples of what is entered for a LaTeX bibliographic reference are:

```
@BOOK{KN:mybook,  
  AUTHOR = "E. A. Poe",  
  TITLE = "A Conchologists First Book",  
  PUBLISHER = "Haswell, Barrington, and Haswell",  
  YEAR = "1839"  
}  
  
@ARTICLE{KN:hopcroft-tm,  
  AUTHOR = "John Hopcroft",  
  TITLE = "Turing Machines",  
  JOURNAL = "Scientific American",  
  YEAR = "1984",  
  PAGES = "86-98"  
}
```

The older REFER format encodes the same things slightly differently:

```
%K mybook  
%A E. A. Poe  
%I Haswell, Barrington, and Haswell  
%T A Conchologists First Book  
%D 1839  
  
%A John Hopcroft  
%T Turing Machines  
%J Scientific American  
%D 1984  
%P 86-98  
%K hopcroft-tm
```

Note the following about REFER:

- Fields are not necessarily in the same order
- There is one field per line
- Each line begins with %
- The third character is a blank
- References are separated by an extra newline character.

There are many other types of bibliographic entries (such as dissertations, government documents, technical reports, etc.) and many other fields (such as volume number, place of publication, editors, etc.). However, for this assignment, we make the simplifying assumption that there are only books and articles, and that they have exactly the fields shown above.

Assignment:

Create a LEX program which will convert a file of REFER entries into LaTeX entries. The LaTeX fields should be in the order shown above, but the REFER fields can appear in any order. Therefore you need to collect the fields and output the LaTeX entry only after reaching the empty line (you may assume the last entry is also followed by an empty line). You must infer the entry type from which fields are present.

If all the entries for an article (or a book) are not present then you need only output: "INVALID ENTRY:Missing fields". You should present a similar message if there are too many fields, if an undefined field is specified (e.g. %V), or if any line of the entry is ill-formed.

What to submit:

- 1) a lex program, as described above.
- 2) your input file, which must test all aspects of your code.
- 3) a typescript session showing the program, the data and a run.
- 4) a README file, in which you discuss the limitations of the approach of inferring an entry type by the choice of fields; create some new types of entries to illustrate your discussion (but do not implement them).