

EXPECIFICACIONES DE LA PRÁCTICA

El formato **BibTeX** es una codificación estándar para elaborar catálogos con información bibliográfica sobre artículos científicos, libros, capítulos de libro, etc. Dado que BibTeX es un formato amplio, trabajaremos sobre una simplificación que se detalla a continuación:

- 1) Un archivo BibTeX es un archivo de texto compuesto por uno o más registros, pero no puede estar vacío, por lo que, un archivo sin un sólo registro debe considerarse erróneo.
- 2) El lenguaje BibTeX tiene un conjunto de palabras reservadas que son: @ARTICLE, @BOOK, @PROCEEDINGS, @INPROCEEDINGS, author, title, journal, year, volume, number, publisher, edition, editor, isbn y pages. Debe tenerse en cuenta que BibTeX no es sensible a mayúsculas.
- 3) Un registro BibTeX está compuesto por una indicación del tipo de registro seguida de la información propiamente dicha entre llaves. El formato es el siguiente:

```
@TIPO-REGISTRO{información del registro}
```

- 4) El tipo de registro es una palabra reservada con un significado específico. Debe ir en mayúsculas y comenzar siempre por el símbolo de la arroba. En nuestra simplificación del lenguaje sólo puede ser uno de los siguientes:

@ARTICLE: se trata de un artículo publicado en una revista científica

@BOOK: el registro contiene información bibliográfica sobre una monografía

@PROCEEDINGS: registro con información sobre las actas de un congreso científico

@INPROCEEDINGS: información sobre un artículo en las actas de un congreso

- 5) Todos los registros tienen información sobre la publicación a que se hace referencia. Dicha información está dividida en campos separados por comas que se incluyen dentro de las llaves que delimitan la información sobre el registro según el siguiente ejemplo:

```
@TIPO-REGISTRO{identificador,  
nombre_campo1 = {información campo 1},  
nombre_campo2 = {información campo 2},  
nombre_campo3 = {información campo 3}  
}
```

- 6) El identificador de registro es un campo obligatorio que, además debe ser único (no puede haber repetidos). Este identificador aparece inmediatamente después de la llave de apertura y está separado del resto de la información por una coma. Dicho indicador será obligatoriamente una palabra entre 1 y 10 caracteres alfanuméricos que comenzará con letra y que podrá contener única y exclusivamente letras, dígitos y los símbolos "-" y "_".

Así, por ejemplo:

```
Abran2002, Sicilia_08, Garcia-07b y Rodr2009
```

serán identificadores válidos.

Mientras que:

```
AlainAbran2002 (demasiado largo),
```

```
Sicilia#08 (símbolo no permitido) y
```

```
07Garcia (símbolo inicial incorrecto)
```

no lo serán.

Un ejemplo de registro válido sería, por tanto:

```
@ARTICLE{Pnueli98,  
  < resto de la información >  
}
```

7) Dentro de cada campo puede haber cadenas de texto (que pueden contener únicamente mayúsculas, minúsculas, puntos y espacios), números enteros (por ejemplo, las fechas) o rangos de página (dos números separados por dos signos menos "--", p.e. 23--28).

Toda esta información puede aparecer en cualquier orden dentro de las llaves que delimitan la información sobre el registro, excepto el identificador de registro (ver puntos 5 y 6) que es obligatorio y común a todos los tipos de registro y que aparece inmediatamente después de la llave de apertura. Algunas de las informaciones que podrán aparecer son:

```
author = {nombre y apellidos del autor}  
title = {título del artículo}  
journal = {nombre de la revista donde se publicó}  
year = {4 dígitos año de publicación}  
volume = {número de volumen}  
number = {número de la revista}  
publisher = {nombre de la editorial que lo publicó}  
edition = {número de edición}  
isbn = {código ISBN, formado por 10 dígitos}  
pages = {página inicial - página final}  
month = {mes en que se publicó}  
editor = {nombre de los editores}
```

No obstante, como ya se ha dicho, no en todos los registros aparecen todas pues unas son aplicables a un tipo de registro, pero no a otro. Así, el campo `ISBN` no es aplicable a los registros de tipo `@ARTICLE`, aunque esto no deberá comprobarlo como parte de la práctica.

8) En el caso de registros de tipo artículo (@ARTICLE), además del identificador es obligatorio consignar información sobre el autor del mismo, el año de publicación, la revista en que se publicó y el título del artículo. La información para dichos campos irá en el orden que se indica a continuación, y cumplirá el siguiente formato y orden:

```
author = {Nombre y apellidos del autor.  
         Si hay más de uno, separados por la palabra "and"}  
title = {título del artículo}  
journal = {nombre de la revista donde se publicó}  
year = {cuatro dígitos para el año de publicación}
```

Cada campo de información debe ir separado de los demás por comas, excepto el que va inmediatamente antes de la llave de cierre, según el siguiente ejemplo:

```
@ARTICLE{Pnueli98,  
  author = {A. Pnueli and O. Shtrichman and M. Siegel},  
  title = {The code validation tool},  
  journal = {Software Tools for Technology Transfer},  
  year = {1998}  
}
```

Otra información que puede reseñarse opcionalmente en los artículos es el número de volumen y el número de la revista. El número de volumen será cualquier entero positivo, y el número de la revista un entero entre 1 y 12. El formato para esta información es el siguiente:

```
volume = {número de volumen}  
number = {número de la revista}
```

Obsérvese que, si aparecen ambas informaciones, el orden será el indicado arriba: primero la información de volumen y luego la del número de la revista.

9) En el caso de los libros (@BOOK), además del identificador es obligatorio consignar información sobre el autor, el título del libro, la editorial y su año de publicación según el siguiente formato (se debe respetar el orden en que aparecen aquí listados):

```
author = { Nombre y apellidos del autor.  
         Si hay más de uno, separados por la palabra "and"}  
title = {título del libro}  
publisher = {nombre de la editorial que lo publicó}  
year = {cuatro dígitos para el año de publicación}
```

Otra información que puede reseñarse opcionalmente en los libros es el número de edición y el ISBN, siendo el formato para esta información (de nuevo debe respetarse el orden):

```
edition = {número de edición, que será un entero positivo}  
isbn = {código ISBN, formado exactamente por 10 dígitos}
```

Un ejemplo de registro completo para un libro sería el siguiente:

```
@BOOK{grune_2000,  
  author = {D. Grune and C. Jacobs and K. Langen and H. Bal},
```

```

    title = {Modern Compiler Design},
    publisher = {John Wiley and Sons, Inc. New York, NY, USA},
    year = {2000},
    edition = {1},
    isbn = {0471976970}
}

```

No obstante, como ya se ha dicho, no en todos los registros aparecen todas pues unas son aplicables a un tipo de registro, pero no a otro. Así, el campo ISBN no es aplicable a los registros de tipo @ARTICLE, aunque esto no deberá comprobarlo como parte de la práctica.

10) En el caso de los artículos dentro de un libro de actas (@INPROCEEDINGS), además del identificador es obligatorio consignar información sobre el autor del mismo, el año de publicación, el nombre del libro de actas en que aparece y el título del artículo (respetando dicho orden):

```

author = { Nombre y apellidos del autor.
          Si hay más de uno, separados por la palabra "and"}
year = {cuatro dígitos para el año de publicación}
title = {título del artículo}
booktitle = {nombre del libro de actas donde se publicó}

```

Otra información que puede reseñarse opcionalmente en los artículos es el nombre del editor, el mes en que se celebró el congreso o conferencia que dio lugar al libro de actas y cualesquiera notas aclaratorias. El formato para esta información es:

```

month = {un texto libre con el mes}
editor = {nombre de los editores separados por "and"}
note = {texto libre con notas aclaratorias}

```

Un ejemplo sería el siguiente:

```

@INPROCEEDINGS{Barnett-Schulte01a,
  author = {Mike Barnett and Wolfram Schulte},
  year = {2001},
  title = {Spying on Components},
  booktitle = {Workshop on Specification Systems},
  month = {Oct},
  editor = {Gary Leavens and M. Sitaraman and D. Johnson},
  note = {Published as Iowa State Technical Report 01-09a}
}

```

11) Finalmente, en las actas de congresos (@PROCEEDINGS) sólo es obligatorio consignar el título y el año (en este orden):

```

@PROCEEDINGS{IFIP95,
  title = {3rd conference on compilation techniques},
  year = {1995},
}

```

Otra información que puede reseñarse opcionalmente es el nombre del editor o editores (si son varios separados por la palabra "and") y el ISBN, siendo el formato para esta información:

```
editor = {nombre de los editores separados por "and"}  
isbn = {código ISBN, formado exactamente por 10 dígitos}
```

Algunos ejemplos de registros correctos serían:

```
@ARTICLE{Pnueli98,  
  author = {A. Pnueli and O. Shtrichman and M. Siegel},  
  title = {The code validation tool CVT},  
  journal = {Software Tools for Technology Transfer},  
  pages = {23--45},  
  year = {1998}  
}  
@BOOK{grune_2000,  
  title = {Modern Compiler Design},  
  isbn = {0471976970},  
  author = {D. Grune and C. Jacobs and K. Langendoen and H. Bal},  
  year = {2000},  
  edition = {1}  
}  
@PROCEEDINGS{IFIP95,  
  title = {Proc. of the 3rd conference on compilation techniques},  
  year = {1995},  
  editor = {John P. Lucas}  
}
```

NOTA: Debe tenerse en cuenta que un archivo BibText deberá obligatoriamente contener un registro (o más) para ser considerado válido.

PRÁCTICA

A.- Se pide programar un analizador léxico, sintáctico y semántico, utilizando JFlex y CUP, que sea capaz de procesar archivos BibTex y emitir informes con la siguiente información sobre errores:

- Claves de registro incorrectas: supera la longitud máxima, tiene un carácter incorrecto, etc.
- Existencia de dos o más entradas con clave repetida
- Errores en fechas (en campos `year`, ha de comprobarse que la fecha tiene 4 dígitos, en `month`, que esté entre 1 y 12, etc.)
- Errores en el `ISBN`, que ha de estar formado por 10 dígitos
- Errores léxicos derivados de la existencia de símbolos no permitidos
- Números de página incorrectos (la página inicial es mayor que la final, p.e. 45--23)

Comprobar algunos errores comunes, tales como:

- Que en la lista de autores de una publicación no aparezca ningún autor repetido.
- Que en la lista de editores no aparezca ningún editor repetido.
- Que ningún autor de una publicación es a su vez editor de la misma.
- Que las publicaciones anteriores a 1920 no tienen ISBN.

B.- Su analizador deberá comprobar que la extensión del archivo es `.bib` y que no se encuentra vacío.

C.- El nombre del analizador léxico debe ser `BibtexScan`.

D.- Se valorará el uso justificado de estados léxicos en la resolución de la práctica.

E.- Mostrar estadísticas al final de la ejecución:

- Número promedio de autores por artículo.
- Autor con mayor número de publicaciones.
- Título de la publicación más antigua.
- Título de la publicación más reciente.

F.- Comprobar la existencia de publicaciones de muy alto impacto. Esto debe comprobarse sólo para los artículos (`@ARTICLE`) y se sabrá si la publicación es o no de alto impacto si se ha publicado en alguno de las siguientes revistas:

- Journal of the Royal Society Interface
- Science
- Nature
- Naturwissenschaften
- Journal of the New York Academy of Sciences
- Scientific American
- Experientia
- Philosophical Transactions of the Royal Society of London
- Proceedings of the National Academy of Sciences
- IBM Journal of Research and Development

Para estos casos, debe mostrarse un resumen, por pantalla, del título del artículo (`@ARTICLE`), su fecha y autores, e indicar que se trata de una publicación de alto impacto.

REFERENCIAS

[Bibtex]

<http://www.bibtex.org/>

EVALUACIÓN DE LAS PRÁCTICAS

Elementos que se esperan en el envío de la práctica:

- Ficheros fuentes de la implementación. (JFlex y CUP)
- Al menos, 8 casos de prueba, 4 correctos y 4 erróneos.
- Memoria de la práctica.

El software entregado debe estar libre de virus, en caso de encontrar alguna práctica con virus se considerará no entregada y por lo tanto suspensa.

Calificación

Se evaluará:

- a) Funcional 20% - Hace lo que se pide (cumple los todos los puntos del análisis léxico, sintáctico y semántico)
- b) Estructura Gramática 30% La gramática no tiene problemas de diseño (desplazamiento/reducción o reducción/reducción)
- c) Arquitectura Solución 20% Se hace utilizando estructuras adecuadas de JAVA (map, hashmap, ...)
- d) Aspectos Formales 30% Se explica adecuadamente en la memoria y se demuestra su funcionamiento con los casos de prueba

La calificación de la práctica se divide en los siguientes niveles:

aprobado (5): Implementación de la especificación léxica y sintáctica (completa).

Al producirse un error debe mostrarse la línea y columna del error.

aprobado (6): Implementación de la especificación léxica y sintáctica.

Se ha sobre-escrito el tratamiento de errores.

notable (7-8,5): Implementación al completo la especificación semántica

Se otorgará un **1,5 extra** en función de la calidad de la memoria entregada.

Se podrá pedir que los alumnos defiendan la práctica presencialmente, si lo considera necesario el profesor.

Si se detecta alguna práctica copiada, los alumnos afectados tendrán la práctica y la asignatura suspensa en la presente convocatoria.