

Online Databases in SAGE

Accessing the Sloane Online Encyclopedia of
Integer Sequences

Steven Sivek
Massachusetts Institute of Technology

Sloane Online Encyclopedia of Integer Sequences

- Searchable database with over 110000 sequences of integers
- For each sequence: integer values, English description, formula, references, etc.
- Only accessible through web interface
- Goal: create SAGE interface for OEIS

OEIS interface: Lookup

- Simple version: look up sequence by number
- Submit queries to OEIS website, read results between `<PRE>` tags, return them

```
f = urllib.urlopen("http://www.research.att.com/.../eisA2.cgi?Anum=A%s"%number)
s = f.read()
i, j = s.find("<PRE>"), s.find("</PRE>")
return parse_sequence(s[i+5:j].strip())
```

- Parsing is easy using the `re` library: all lines have the form `"%L Annnnnnn data"`

OEIS interface: Parsing results

- Sample text:

```
%S A000001 1,1,1,2,1,2,1,5,2,2,1,5,1,2,1,14,...  
%N A000001 Number of groups of order n.
```

- Determine type of line, sequence, and data:

```
entry = re.compile(r'%(?P<letter>[A-Za-z]) '  
    r'A(?P<number>\d{6}) (?P<body>.* )$')  
m = entry.search(line)  
if m.group('letter') == 'S':  
    list += m.group('body').split(',')  
elif m.group('letter') == 'N':  
    description = m.group('body').strip()
```

OEIS interface: Lookup

- Example: Busy Beaver sequence, A060843

```
sage: sloane_sequence(60843)
```

```
[60843, 'Busy Beaver problem: maximal number  
of steps that an n-state Turing machine can  
make on an initially blank tape before eventually  
halting.', [1, 6, 21, 107]]
```

- Results have the form
[number, description, list].

OEIS interface: Searching

- It is more useful to look up sequences by consecutive terms than by OEIS number
- Web interface: pass comma-separated list, get sequences between <PRE> tags
- Results are separated by blank lines and can be parsed by the same code as before

```
for seq in text.split('\n\n'):  
    if seq.find('%') != -1:  
        results.append(parse_sequence(seq))
```

Problems with the interface

- The OEIS server sometimes returns errors

```
sage: sloane_sequence(60843)
```

```
[]
```

```
sage: sloane_sequence(60843)
```

```
[60843, ...]
```

- OEIS searches are limited to 100 results
- The OEIS server is *slow*: even simple searches can take up to 30 seconds

Offline interface

- The solution is to create an offline version
- OEIS provides a list of all the sequences which is 6.4 MB when bziped, so it is included as an optional package

```
sage -i database_sloane_oeis-2005-12
```
- The file is easily searchable, but provides only the actual sequences (no descriptions)

Offline interface: Design

- Create a class with necessary methods, but only make a single instance available.
- Treat the instance as an array of sequences.

```
sloane.py: Class SloaneEncyclopediaClass:  
            def __getitem__(self, N):  
                ...  
            SloaneEncyclopedia = SloaneEncyclopediaClass()
```

```
all.py:      from sloane import SloaneEncyclopedia
```

```
sage: SloaneEncyclopedia[60843]  
[1, 6, 21, 107]
```

Offline interface: Design

- Each sequence is stored simultaneously in two forms: a list of integers and a string

```
sage: SloaneEncyclopedia.__data__[60843]
[60843, [1, 6, 21, 107], '1,6,21,107,']
```
- Lists: fast for iterating through all sequences

```
sage: diag = [ SloaneEncyclopedia[i][i]+1
sage: diag = [ for i in range(1,42) ]
```
- Strings: fast for performing searches

```
sage: time len(SE.find([1,2,3,4,5], 10000)
2661
Time: CPU 1.03 s, Wall: 1.17 s
```

Offline interface: Memory

- Storing the database in two forms consumes about 80 MB, and initializing it takes 20 s.
- Keep `self.__loaded__` flag, and only load the database when `__loaded__` is false and the database is accessed.
- If the user needs more available memory, `SloaneEncyclopedia.unload()` simply deletes `self.__data__`.

Comparison of the designs

Online:

- More informative (descriptions)
- Consumes very little memory, CPU time
- Frequently updated

Offline:

- More stable than OEIS server
- Unlimited search results
- Much faster

In practice, both are useful: the offline one to find sequences, the online one to learn what they are.