

The Approach:

The entire design is divided into three stages.

1. Parsing Input GEDCOM file. In this stage, [FileParser](#) -> [GedcomFileParser](#) will parse the GEDCOM file and will return a list of [Entity](#) objects. Each line of input record is converted into [Record](#) object and all logically related records belong to one [Entity](#). Also, [Record](#) object is responsible how to split each record.

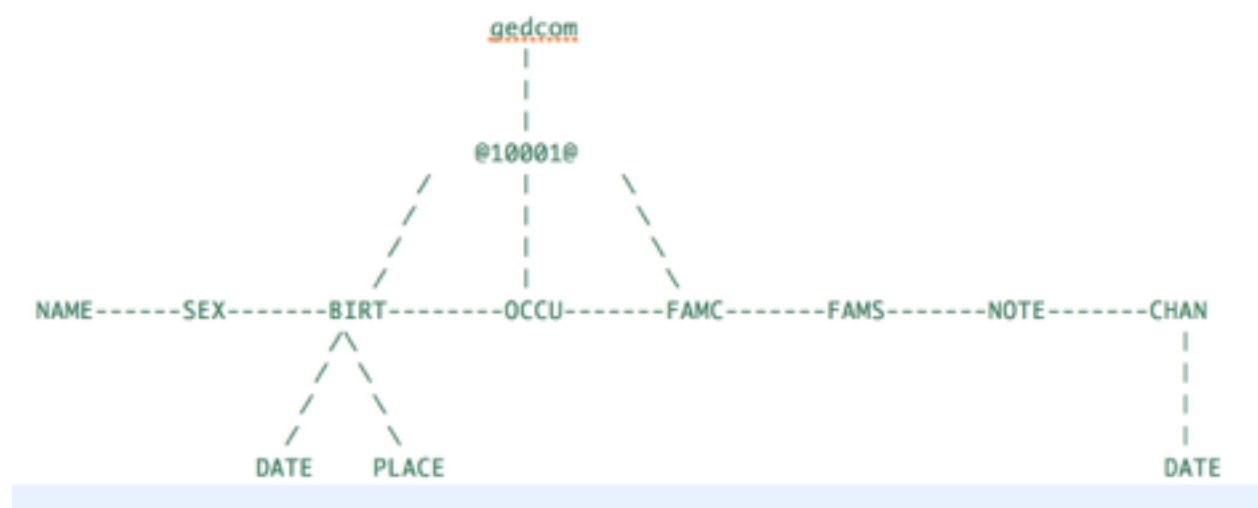
Advantage -

1. Easy to count how many entities are in GEDCOM file.
2. Easy to count how many records belong to one entity.
3. Highly cohesive objects with single responsibility.
4. Unit testable.

The below record will get converted into one entity object with 12 records.

```
0 @I0001@ INDI
1 NAME Elizabeth Alexandra Mary /Windsor/
1 SEX F
1 BIRT
2 DATE 21 Apr 1926
2 PLAC 17 Bruton Street, London, W1
1 OCCU Queen
1 FAMC @F0003@
1 FAMS @F0001@
1 NOTE @N0002@
1 CHAN
2 DATE 13 Dec 2003
```

2. Building Intermediate state i.e. [EntityTree](#). List of [Entity](#) objects will be passed to [TreeBuilder](#) -> [GedcomTreeBuilder](#) which will build an [EntityTree](#) and create [Node](#) object for each [Record](#) along with their hierarchy information, as shown below.



GEDCOM Parser Challenge

Advantage -

1. Having an intermediate state of EntityTree will decoupled input format of GEDCOM file with output format of XML.
2. EntityTree contains level, parent and children for each node, so easy to identify where the node belong in tree structure.
3. Bi-directional support for traversing tree, as each node has information of their parent and children.
4. Easy to identify root and leaf nodes.
5. Unit testable.

3. In last stage, [EntityTree](#) is passed to [FileBuilder](#) -> [XmlBuilder](#) which takes [EntityTree](#) and build an XML as shown below:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<gedcom>
  <INDI ID="@I0001@">
    <NAME>Elizabeth Alexandra Mary /Windsor/</NAME>
    <SEX>F</SEX>
    <BIRT>
      <DATE>21 Apr 1926</DATE>
      <PLAC>17 Bruton Street, London, W1</PLAC>
    </BIRT>
    <OCCU>Queen</OCCU>
    <FAMC ID="@F0003@" />
    <FAMS ID="@F0001@" />
    <NOTE ID="@N0002@" />
    <CHAN>
      <DATE>13 Dec 2003</DATE>
    </CHAN>
  </INDI>
```

Advantage -

1. Incase, if need to support multiple output formats then the same EntityTree can be passed.
2. Unit testable.

Assumptions:

1. Record are split using one or more whitespace.
2. First level of hierarchy starts with 0.
3. ID starts with @.

Note - [Assumptions.Java](#) - list all assumptions.

GEDCOM Parser Challenge