1. [10 points] Choose a document or piece of text that you're interested in from your workplace. This document can be structured or unstructured. You may choose a text of any sort and any length, as long as it is long enough to meet the following encoding criteria. Be sure to include the text as a separate file in your upload (or if it is online, you may provide a link to it)

2. [20 points] Make an XML DTD for that text. Your DTD should specify the following:

- At least 10 different elements
- At least 5 different attributes (you may have multiple attributes per element; not every element requires attributes). For at least 2 attributes, a controlled list of values the attribute may take.
- Indicate, of course, whether each element and attribute is required, optional, an either/or, repeated, etc.
- Indicate how elements may nest.

3. [10 points] Write prose documentation for each element, attribute, and attribute value.

4. [10 points] Mark up the text you have chosen according to the DTD you designed. Make sure you include either a <!DOCTYPE> reference to an external DTD file, or an internal DTD within the <! DOCTYPE> element in the XML document. Be sure to validate your document (see below) as validation is one of the areas you will be assessed on. Be sure to validate your document (see below) as validation is one of the areas you will be assessed on.

5. [25 points] Write a narrative about this process, answering the following reflection questions:

- How did you decide to represent the data in the way that you did? Why did you choose the elements and attributes that you did?
- What were the hardest decisions you had to make in this design process?
- How does your DTD design support data independence?
- How may your DTD design support the overarching goals of data curation (revisit objectives and activities of Week 1)?
- What are the pros and cons of your DTD design?

6. Each student will be required to grade the submissions of 5 of their peers. Submissions will be graded based on the following criteria. Write a constructive and professional review and post to the course forum replying to the individual's submission.
a. Is everything represented?
b. Is it clearly written? Is the scheme and data clearly presented?
c. What are the pros and cons of this representation?

d. How could it be done differently? How could it be improved?

Assignment 2 – XML Schema Design Exercise
- RAJA SHEKAR BOLLAM (bollamr2)

For Assignment 2, I choose to create a DTD document for my Resume. I would like to create XML Elements for each sections and details from my resume document.
Below are the documents that are submitted for the assignment purpose.
- DTD Files
- DTD + XML File Integrated
- Sample Document (Resume)
- Generated XML output
Note: As per the peer review comments, document and DTD is updated after the initial submission.

Explanation of DTD Elements
- resume, personal, introduction, education, professionalexperience, technicalskills, others

Resume
DocumentType **resume** comprises of sections like personal, introduction, education, profession experience, technicalskills and others. Below is the explanation for each sections of the DTD.
dtd snapshot:
```
<!ELEMENT resume (personal, introduction, education, professionexperience+, technicalskills, others*)>
```

Personal
This section has 3 attributes Name, Phone number and Email ID grouped together as personalInfo. I have marked all these attributes values a REQUIRED. It was intentional to mark each attribute as Elements to understand the shortcomings of it. Also I have empty initialized all the attributes. Note that in case of email-Id and phone-number, there could be more than one instance of the element.
dtd snapshot:
```
<!ELEMENT personal (personalinfo)>
<!ELEMENT personalinfo (personalinfoName, personalinfoPhone+, personalinfoEmail+)>
<!ELEMENT personalinfoName EMPTY>
<!ATTLIST personalinfoName name CDATA #REQUIRED>
<!ELEMENT personalinfoPhone EMPTY>
<!ATTLIST personalinfoPhone phonenumber CDATA #REQUIRED>
<!ELEMENT personalinfoEmail EMPTY>
<!ATTLIST personalinfoEmail email CDATA #REQUIRED>
```

Introduction
This section is an introductory note of the person. It is more of a summary of the resumes owners. This section is categorized as ANY, meaning it contain any combination of parsable data.
dtd snapshot:
```
<!ELEMENT introduction ANY>
```

Education
This section is intentionally made as plain as simple. It has at least **ONE** element institute that constitutes of instinstitute Name, Type, Year of Starting, Year of Ending and Major. Each item is an parsable element in itself.

dtd snapshot:
```
<!ELEMENT education (institute+)>
<!ELEMENT institute (instituteName, instituteType, instituteYearStart, instituteYearEnd, instituteMajor)>
```

ProfessionalExperience

This section details the experience in related to profession. This section is grouped as atleast one instance of prefessioinalExperience. Each professionalExperience is inturn grouped as “employer”, “YearsInThisCompany” and more than one instance of “Projects”. Each Project is inturn grouped as “ProjectName”, “Designation”, “YearStart”, “YearEnd”, “Duration” and “Description”.

Note that YearStart, YearEnd, Duration can exist in more than one combination as the meaning combination is needed in different scenarios. For example, Incase of 2 projects in one year, YearStart and Duration can be used.

dtd snapshot:

```
<!ELEMENT professionexperience (professionalExperience+)>
<!ELEMENT professionalExperience (Employer, YearsInThisCompany, Projects+)>
<!ELEMENT Employer (#PCDATA)>
<!ELEMENT YearsInThisCompany (#PCDATA)>
<!ELEMENT Projects (ProjectName, Designation, YearStart?, YearEnd?, Duration?, Description)>
<!ELEMENT ProjectName (#PCDATA)>
<!ELEMENT Designation (#PCDATA)>
<!ELEMENT YearStart (#PCDATA)>
<!ELEMENT YearEnd (#PCDATA)>
<!ELEMENT Duration (#PCDATA)>
<!ELEMENT Description (#PCDATA)>
```

TechnicalSkills

This section we group the technical capabilities of the candidate. In this section we use the proper way of using a attributes unlike from the section *“personal”*. *Domain, Languages, skills, protocols, databases and Operating System* information is grouped together in this section. Notice that we have marked “databases” and “OS” as *“IMPLIED”* attribute type. To explain about default initialization we have default values for skills.

dtd snapshot:

```
<!ELEMENT technicalskills (skills)>
<!ELEMENT skills (#PCDATA)>
<!ATTLIST skills
domain CDATA "Mobile"
languagues CDATA "Computer Programming"
skills CDATA "Computer Frameworks"
procotols CDATA "TCP/IP"
databases CDATA #IMPLIED
OS CDATA #IMPLIED
>
```

Others

This section is to have any miscellaneous information that needs to be presented. Notice that this section is marked with (*). This is an optional element.

dtd snapshot:

```
<!ELEMENT others (#PCDATA)>
```

DTD for reference purpose: *(This contents are same as DTD_assignment2_bollamr2.txt)*

<?xml version="1.0" encoding="UTF-8"?>

```
<!DOCTYPE resume [
<!ELEMENT resume (personal, introduction, education, professionexperience+, technicalskills, others*)>
<!ELEMENT personal (personalinfo)>
<!ELEMENT personalinfo (personalinfoName, personalinfoPhone+, personalinfoEmail+)>
<!ELEMENT personalinfoName EMPTY>
<!ATTLIST personalinfoName name CDATA #REQUIRED>
<!ELEMENT personalinfoPhone EMPTY>
<!ATTLIST personalinfoPhone phonenumber CDATA #REQUIRED>
<!ELEMENT personalinfoEmail EMPTY>
<!ATTLIST personalinfoEmail email CDATA #REQUIRED>
<!ELEMENT introduction ANY>
<!ELEMENT education (institute+)>
<!ELEMENT institute (instituteName, instituteType, instituteYearStart, instituteYearEnd, instituteMajor)>
<!ELEMENT instituteName (#PCDATA)>
<!ELEMENT instituteType (#PCDATA)>
<!ELEMENT instituteYearStart (#PCDATA)>
<!ELEMENT instituteYearEnd (#PCDATA)>
<!ELEMENT instituteMajor (#PCDATA)>
<!ELEMENT professionexperience (professionalExperience+)>
<!ELEMENT professionalExperience (Employer, YearsInThisCompany, Projects+)>
<!ELEMENT Employer (#PCDATA)>
<!ELEMENT YearsInThisCompany (#PCDATA)>
<!ELEMENT Projects (ProjectName, Designation, YearStart?, YearEnd?, Duration?, Description)>
<!ELEMENT ProjectName (#PCDATA)>
<!ELEMENT Designation (#PCDATA)>
<!ELEMENT YearStart (#PCDATA)>
<!ELEMENT YearEnd (#PCDATA)>
<!ELEMENT Duration (#PCDATA)>
<!ELEMENT Description (#PCDATA)>
<!ELEMENT technicalskills (skills)>
<!ELEMENT skills (#PCDATA)>
<!ATTLIST skills
domain CDATA "Mobile"
languagues CDATA "Computer Programming"
skills CDATA "Computer Frameworks"
procotols CDATA "TCP/IP"
databases CDATA #IMPLIED
OS CDATA #IMPLIED
>
<!ELEMENT others (#PCDATA)>
]>
```

Review Criteria and Points Distribution:

1. For this assignment purpose, I choose Resume Builder for XML DTD development. As a reference and testing I used my own resume.
2. Atleast 10 different Elements, 5 different attributes, 2 Attributes controlled list of values are made.
3. Prose and Cons are listed while describing each ELEMENT of the DTD Schema.
4. Narrative About the process is as explained below:
    1. **How did you decide to represent the data in the way that you did? Why did you choose the elements and attributes that you did?**
    - I used tabular-hierarchical representation of information that can be organized in a resume. As result, I choose personal, introduction, education, professionalexperience, technicalskills and others as tabular hierarchies.
    2. **What were the hardest decisions you had to make in this design process?**
    - Deciding the document/reference document for the DTD design was the most difficult decision to take.
    - Organizing the data to best describe the professional experience was another difficult task.
    3. **How does your DTD design support data independence?**
    - The tabular-hierarchical structure is very well designed in a way that there is no overlap of information between the elements. This can be notice while inspecting the DTD schema in detail.
    4. **How may your DTD design support the overarching goals of data curation (revisit objectives and activities of Week 1)?**
    - DTD schema is organized efficiently in a way to have better ease of understanding the resume, easily extend the elements or section of the DTD with limited changes, with no duplicacy of elements or data.
    - Activities like Collection (support and acquisition of data), Organization (employed appropriate data model and use appropriate standards), Storage(support reliable and effective storage in the form of tabular-hierarchical structure), Preservation(data is understandable and reusable), Discover-ability (ability to search for and located relevant data), Sharing(DTD schema and data can be shared and easily understood), Modification (support management of corrections and updates) have been addressed.
    - Reproducibility activity is supported as a generic validator but not as an element-wise validator.
    - Activities like Access(ability to retireve and distribute data), Workflow(ability to systematize work with data), , Identification (ability to identify, authenticate and validate data), , Integration(integration of data from different sources using different models), , Reformatting (reformatting for use by different tools or to match new format standards), Reproducible (reproduce results ensuring scientific validity and reliability), Security, Compliance, Provenance (support identifying what inputs, calculations and actions are responsible for data values), Communication (support representation, publishing and visualization that could provide insight) have not been addressed in this assignment.
    5. **What are the pros and cons of your DTD design?**
    Pros:
        - This DTD design is a good way of filtering our the resumes in a pre-defined procedure.

- My design is well structure, well abstracted and easy to understand.
- Information on demand can be easily retrieved using my design because of the tabular-hierarchical structure of the DTD.

Cons:

-Not may restrictions can be applied in the DTD design for the data format.
- For Instance, the phone number could have be formatted in a specific way. PersonalInfoName could have followed a pattern like (Family Name, Middle Name, Firstname) etc.
- Reformatting activity could have be supported by using some RFC or Technology standards for formatting data like phonenumber, address etc.
- Reproducibility activity could have been supported by providing more specific validators for elements.
- Compliance activity could have been supported by adding and disclaimer element in the DTD.