json23plet

final presentation

presented by:

Yogev Henig

Yehonatan Buchnik

Yaakov Sokolik

under supervision of:

Dr. Oren Mishali

TDK lab

Overview

json23plet is a Linux command line tool, which was developed in the TDK lab as part of the JBS (Jewish Bookshelf) project.

The goal of the JBS project is to build a semantic web database of Jewish books.

The main purpose of json23plet is to provide a simple functionality to deal with any kind of data contained in Json files. json23plet creates RDF statements in Turtle format that help construct a semantic web domain.

JBS workflow

jbs-tagger

jbs-text2json

Json files that contain data

jbs-json23plet

Semantic Web database of the Jewish Bookshelf !

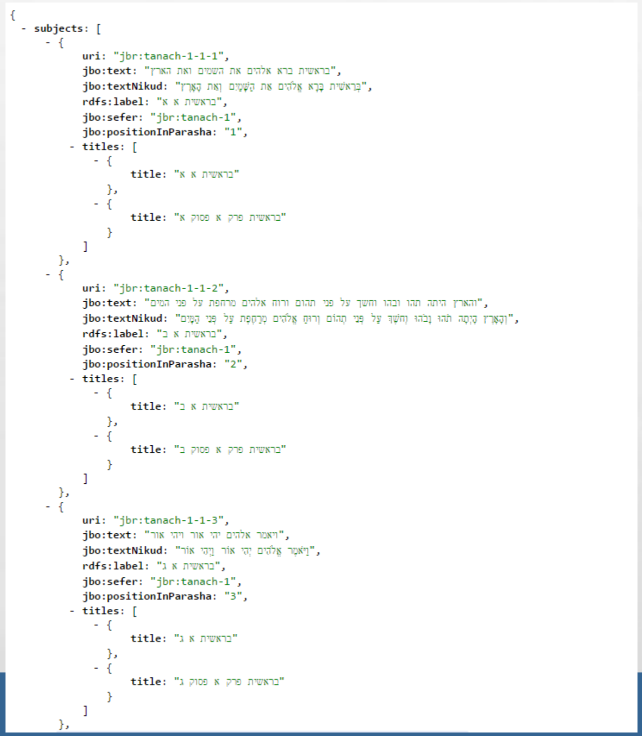
Motivation

- The Jewish book shelf contains a lot of data – many thousands of books.
- Creating a Semantic web directly from it, is a very difficult mission because:
  - There is not a unified format for the text
  - Ontology may change
  - It is hard to identify relations
  - And more…

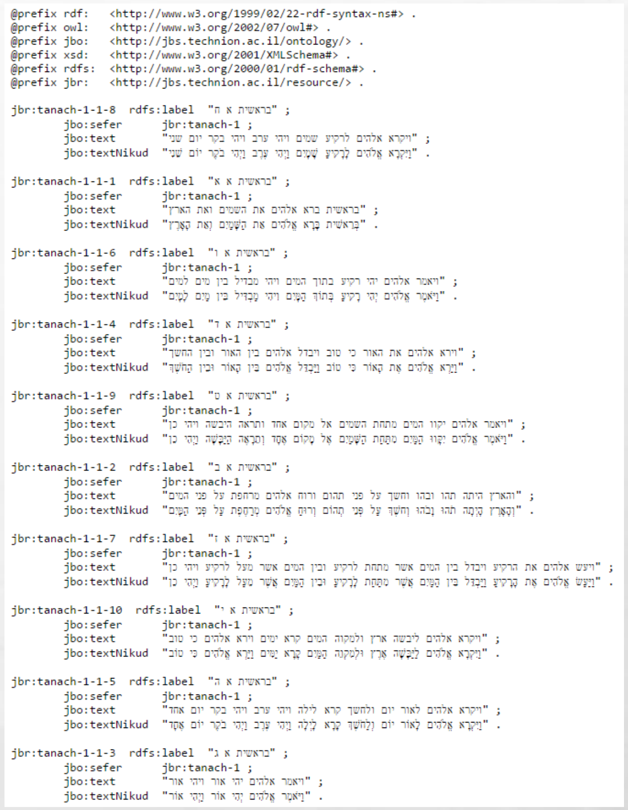
Motivation

- On this task different teams worked parallel:
  - Text2json – create a Json files from a raw text
  - Taggers – identify relations inside the raw text and save it in Json files
  - Json23plet – create RDF model from all Json files.
- This method of work achieved:
  - Parsing raw text only once.
  - Regenerate RDF triples from Json files easily due to their simple format.
  - Maintain and development become easy.

From raw data to RDF



```
{
  - subjects: [
      - {
           uri: "jbr:tanach-1-1-1",
           jbo:text: "בראשית ברא אלהים את השמים ואת הארץ",
           jbo:textNikud: "בְּרֵאשִׁית בָּרָא אֱלֹהִים אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ",
           rdfs:label: "בראשית א א",
           jbo:sefer: "jbr:tanach-1",
           jbo:positionInParasha: "1",
         - titles: [
             - {
                   title: "בראשית א א"
               },
             - {
                   title: "בראשית פרק א פסוק א"
               }
           ]
        },
      - {
           uri: "jbr:tanach-1-1-2",
           jbo:text: "והארץ היתה תהו ובהו וחשך על פני תהום ורוח אלהים מרחפת על פני המים",
           jbo:textNikud: "וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וְחֹשֶׁךְ עַל פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל פְּנֵי הַמָּיִם",
           rdfs:label: "בראשית א ב",
           jbo:sefer: "jbr:tanach-1",
           jbo:positionInParasha: "2",
         - titles: [
             - {
                   title: "בראשית א ב"
               },
             - {
                   title: "בראשית פרק א פסוק ב"
               }
           ]
        },
      - {
           uri: "jbr:tanach-1-1-3",
           jbo:text: "ויאמר אלהים יהי אור ויהי אור",
           jbo:textNikud: "וַיֹּאמֶר אֱלֹהִים יְהִי אוֹר וַיְהִי אוֹר",
           rdfs:label: "בראשית א ג",
           jbo:sefer: "jbr:tanach-1",
           jbo:positionInParasha: "3",
         - titles: [
             - {
                   title: "בראשית א ג"
               },
             - {
                   title: "בראשית פרק א פסוק ג"
               }
           ]
        },
```



```
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix jbo:   <http://jbs.technion.ac.il/ontology/> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jbr:   <http://jbs.technion.ac.il/resource/> .

jbr:tanach-1-1-8  rdfs:label  "בראשית א ח" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויקרא אלהים לרקיע שמים ויהי ערב ויהי בקר יום שני" ;
        jbo:textNikud  "וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמָיִם וַיְהִי עֶרֶב וַיְהִי בֹקֶר יוֹם שֵׁנִי" .

jbr:tanach-1-1-1  rdfs:label  "בראשית א א" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "בראשית ברא אלהים את השמים ואת הארץ" ;
        jbo:textNikud  "בְּרֵאשִׁית בָּרָא אֱלֹהִים אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ" .

jbr:tanach-1-1-6  rdfs:label  "בראשית א ו" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויאמר אלהים יהי רקיע בתוך המים ויהי מבדיל בין מים למים" ;
        jbo:textNikud  "וַיֹּאמֶר אֱלֹהִים יְהִי רָקִיעַ בְּתוֹךְ הַמָּיִם וִיהִי מַבְדִּיל בֵּין מַיִם לָמָיִם" .

jbr:tanach-1-1-4  rdfs:label  "בראשית א ד" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "וירא אלהים את האור כי טוב ויבדל אלהים בין האור ובין החשך" ;
        jbo:textNikud  "וַיַּרְא אֱלֹהִים אֶת הָאוֹר כִּי טוֹב וַיַּבְדֵּל אֱלֹהִים בֵּין הָאוֹר וּבֵין הַחֹשֶׁךְ" .

jbr:tanach-1-1-9  rdfs:label  "בראשית א ט" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויאמר אלהים יקוו המים מתחת השמים אל מקום אחד ותראה היבשה ויהי כן" ;
        jbo:textNikud  "וַיֹּאמֶר אֱלֹהִים יִקָּווּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל מָקוֹם אֶחָד וְתֵרָאֶה הַיַּבָּשָׁה וַיְהִי כֵן" .

jbr:tanach-1-1-2  rdfs:label  "בראשית א ב" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "והארץ היתה תהו ובהו וחשך על פני תהום ורוח אלהים מרחפת על פני המים" ;
        jbo:textNikud  "וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וְחֹשֶׁךְ עַל פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל פְּנֵי הַמָּיִם" .

jbr:tanach-1-1-7  rdfs:label  "בראשית א ז" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויעש אלהים את הרקיע ויבדל בין המים אשר מתחת לרקיע ובין המים אשר מעל לרקיע ויהי כן" ;
        jbo:textNikud  "וַיַּעַשׂ אֱלֹהִים אֶת הָרָקִיעַ וַיַּבְדֵּל בֵּין הַמַּיִם אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמַּיִם אֲשֶׁר מֵעַל לָרָקִיעַ וַיְהִי כֵן" .

jbr:tanach-1-1-10  rdfs:label  "בראשית א י" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויקרא אלהים ליבשה ארץ ולמקוה המים קרא ימים וירא אלהים כי טוב" ;
        jbo:textNikud  "וַיִּקְרָא אֱלֹהִים לַיַּבָּשָׁה אֶרֶץ וּלְמִקְוֵה הַמַּיִם קָרָא יַמִּים וַיַּרְא אֱלֹהִים כִּי טוֹב" .

jbr:tanach-1-1-5  rdfs:label  "בראשית א ה" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויקרא אלהים לאור יום ולחשך קרא לילה ויהי ערב ויהי בקר יום אחד" ;
        jbo:textNikud  "וַיִּקְרָא אֱלֹהִים לָאוֹר יוֹם וְלַחֹשֶׁךְ קָרָא לָיְלָה וַיְהִי עֶרֶב וַיְהִי בֹקֶר יוֹם אֶחָד" .

jbr:tanach-1-1-3  rdfs:label  "בראשית א ג" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "ויאמר אלהים יהי אור ויהי אור" ;
        jbo:textNikud  "וַיֹּאמֶר אֱלֹהִים יְהִי אוֹר וַיְהִי אוֹר" .
```

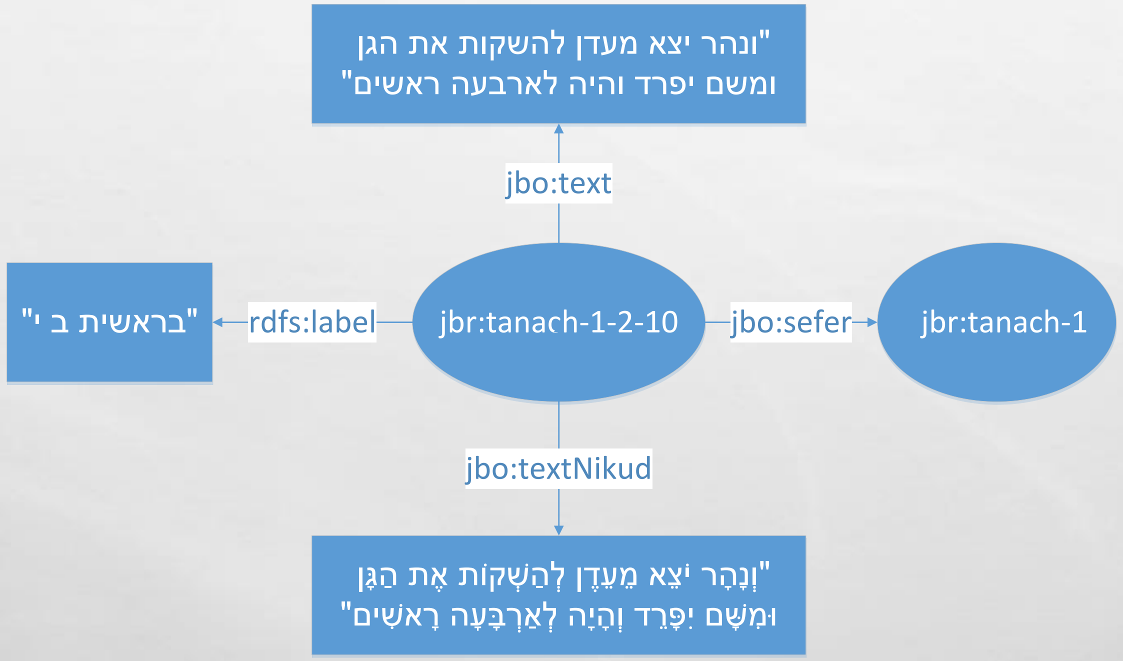
What is ontology?

In order to be able to model entities and abstract data in a domain, we first need to describe a structured common language which gives us the ability to reason over (meta)data.

Creating ontology for the domain allows us to do so.

What are RDF triples?

For example we want to model information of the following RDF diagram:


"ונהר יצא מעדן להשקות את הגן
ומשם יפרד והיה לארבעה ראשים"
jbo:text
"בראשית ב י"
rdfs:label
jbr:tanach-1-2-10
jbo:sefer
jbr:tanach-1
jbo:textNikud
"וְנָהָר יֹצֵא מֵעֵדֶן לְהַשְׁקוֹת אֶת הַגָּן
וּמִשָּׁם יִפָּרֵד וְהָיָה לְאַרְבָּעָה רָאשִׁים"

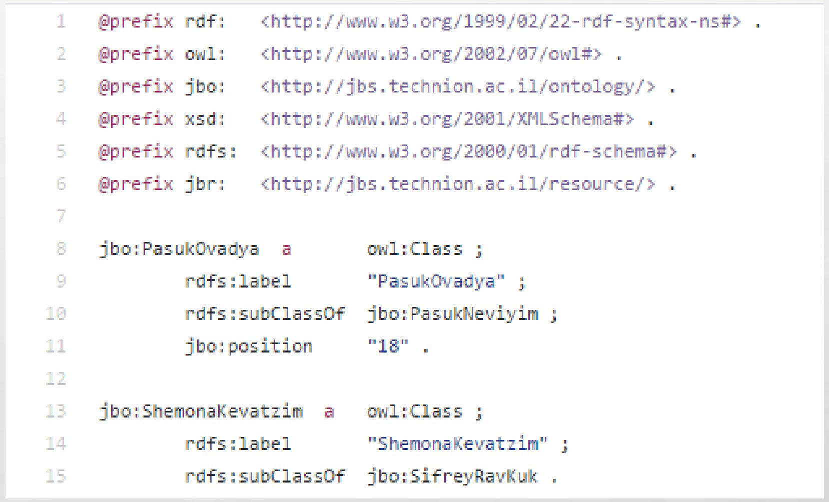
What are RDF triples?

The corresponding triplets are:

subject	predicate	object
jbr:tanach-1-2-10	rdfs:label	"בראשית ב י"
jbr:tanach-1-2-10	jbo:sefer	jbr:tanach-1
jbr:tanach-1-2-10	jbo:text	"ונהר יצא מעדן להשקות את הגן ומשם יפרד והיה לארבעה ראשים"
jbr:tanach-1-2-10	jbo:textNikud	"וְנָהָר יֹצֵא מֵעֵדֶן לְהַשְׁקוֹת אֶת הַגָּן וּמִשָּׁם יִפָּרֵד וְהָיָה לְאַרְבָּעָה רָאשִׁים"

Generating an ontology

- Ontology enable us to describe a structured common language which gives us the ability to reason over (meta)data.
  - Domains
  - Prefixes
  - Classes
  - Predicates
  - And more…

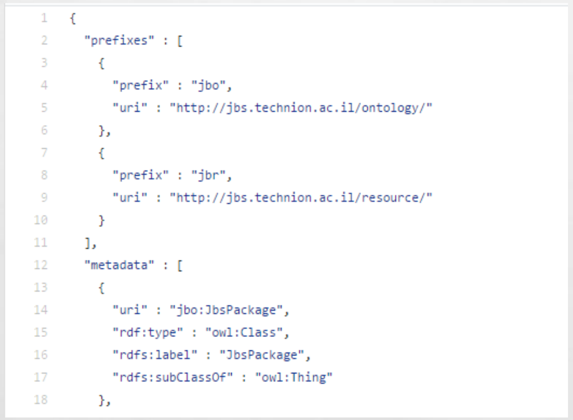


```
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix jbo:   <http://jbs.technion.ac.il/ontology/> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jbr:   <http://jbs.technion.ac.il/resource/> .

jbo:PasukOvadya  a      owl:Class ;
        rdfs:label      "PasukOvadya" ;
        rdfs:subClassOf  jbo:PasukNeviyim ;
        jbo:position     "18" .

jbo:ShemonaKevatzim  a   owl:Class ;
        rdfs:label      "ShemonaKevatzim" ;
        rdfs:subClassOf  jbo:SifreyRavKuk .
```

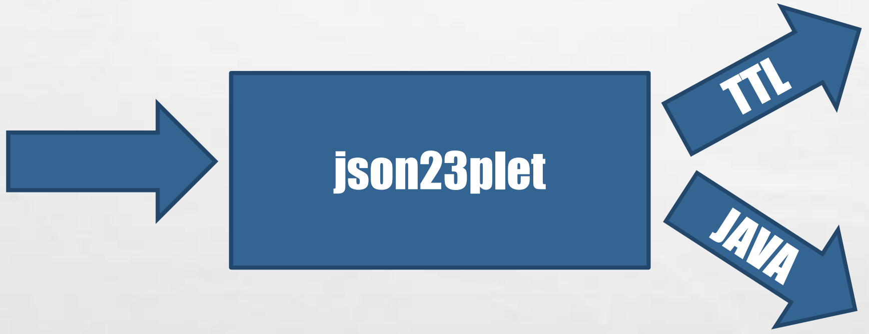
Generating an ontology

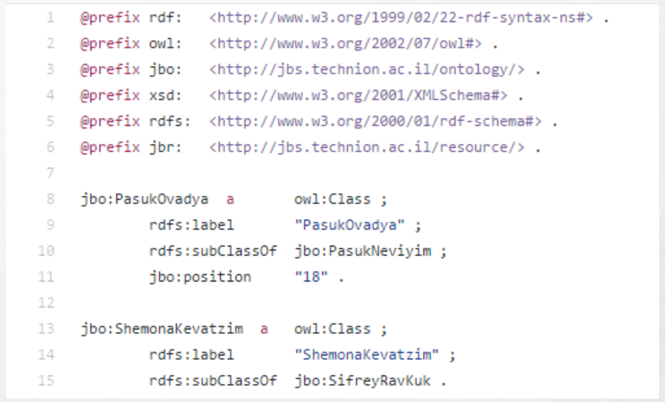
- Json23plet allows you to create, modify and update your ontology in a few simple steps:
  - Create a json file describe the ontology
  - Run json23plet
  - Json23plet will create a ttl file (the ontology) and a java class describe the ontology as java object
    - Useful for referencing to your ontology while writing a generator

Generating an ontology

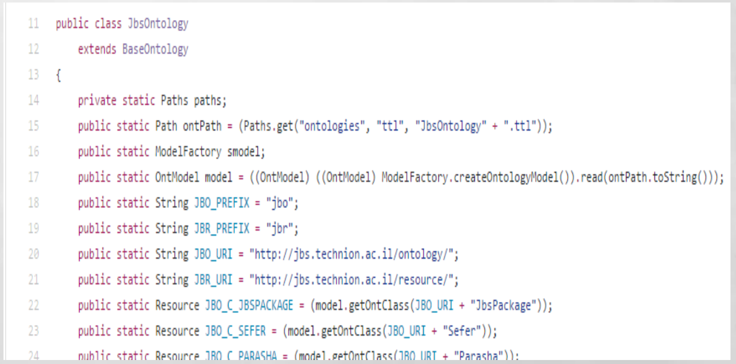


```
{
  "prefixes" : [
    {
      "prefix" : "jbo",
      "uri" : "http://jbs.technion.ac.il/ontology/"
    },
    {
      "prefix" : "jbr",
      "uri" : "http://jbs.technion.ac.il/resource/"
    }
  ],
  "metadata" : [
    {
      "uri" : "jbo:JbsPackage",
      "rdf:type" : "owl:Class",
      "rdfs:label" : "JbsPackage",
      "rdfs:subClassOf" : "owl:Thing"
    },
```





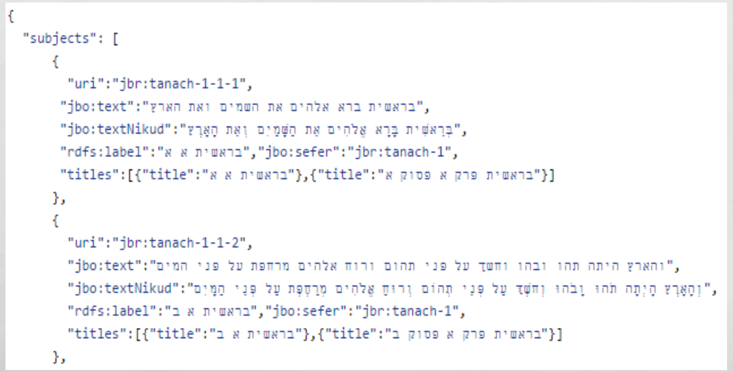
```
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix jbo:   <http://jbs.technion.ac.il/ontology/> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jbr:   <http://jbs.technion.ac.il/resource/> .

jbo:PasukOvadya  a       owl:Class ;
        rdfs:label       "PasukOvadya" ;
        rdfs:subClassOf  jbo:PasukNeviyim ;
        jbo:position     "18" .

jbo:ShemonaKevatzim  a   owl:Class ;
        rdfs:label       "ShemonaKevatzim" ;
        rdfs:subClassOf  jbo:SifreyRavKuk .
```



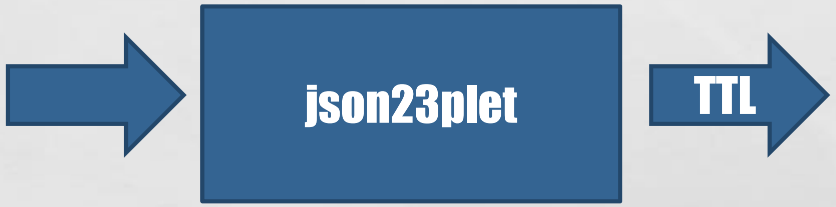
```
public class JbsOntology
    extends BaseOntology
{
    private static Paths paths;
    public static Path ontPath = (Paths.get("ontologies", "ttl", "JbsOntology" + ".ttl"));
    public static ModelFactory smodel;
    public static OntModel model = ((OntModel) ((OntModel) ModelFactory.createOntologyModel()).read(ontPath.toString()));
    public static String JBO_PREFIX = "jbo";
    public static String JBR_PREFIX = "jbr";
    public static String JBO_URI = "http://jbs.technion.ac.il/ontology/";
    public static String JBR_URI = "http://jbs.technion.ac.il/resource/";
    public static Resource JBO_C_JBSPACKAGE = (model.getOntClass(JBO_URI + "JbsPackage"));
    public static Resource JBO_C_SEFER = (model.getOntClass(JBO_URI + "Sefer"));
    public static Resource JBO_C_PARASHA = (model.getOntClass(JBO_URI + "Parasha"));
```

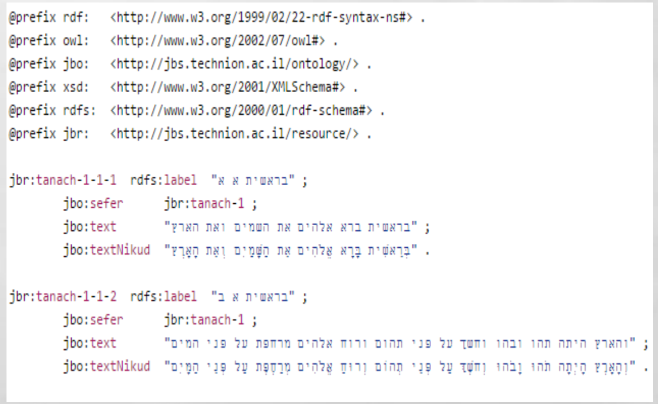
Basic generator

- The basic generator allows you to create a RDF model without any programming skills
- Takes Json files that are in the basic format
- Creats TTL files generically and independently from the semantic of the data



```
{
 "subjects": [
     {
       "uri":"jbr:tanach-1-1-1",
      "jbo:text":"בראשית ברא אלהים את השמים ואת הארץ",
      "jbo:textNikud":"בְּרֵאשִׁית בָּרָא אֱלֹהִים אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ",
      "rdfs:label":"בראשית א א","jbo:sefer":"jbr:tanach-1",
      "titles":[{"title":"בראשית א א"},{"title":"בראשית פרק א פסוק א"}]
     },
     {
       "uri":"jbr:tanach-1-1-2",
       "jbo:text":"והארץ היתה תהו ובהו וחשך על פני תהום ורוח אלהים מרחפת על פני המים",
       "jbo:textNikud":"וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וְחֹשֶׁךְ עַל פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל פְּנֵי הַמָּיִם",
       "rdfs:label":"בראשית א ב","jbo:sefer":"jbr:tanach-1",
       "titles":[{"title":"בראשית א ב"},{"title":"בראשית פרק א פסוק ב"}]
     },
```





```
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix jbo:   <http://jbs.technion.ac.il/ontology/> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix jbr:   <http://jbs.technion.ac.il/resource/> .

jbr:tanach-1-1-1  rdfs:label  "בראשית א א" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "בראשית ברא אלהים את השמים ואת הארץ" ;
        jbo:textNikud  "בְּרֵאשִׁית בָּרָא אֱלֹהִים אֵת הַשָּׁמַיִם וְאֵת הָאָרֶץ" .

jbr:tanach-1-1-2  rdfs:label  "בראשית א ב" ;
        jbo:sefer      jbr:tanach-1 ;
        jbo:text       "והארץ היתה תהו ובהו וחשך על פני תהום ורוח אלהים מרחפת על פני המים" ;
        jbo:textNikud  "וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וְחֹשֶׁךְ עַל פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל פְּנֵי הַמָּיִם" .
```

Tool’s dependencies

We used:

- Apache Jena library to build the RDF model and export it into Turtle files
- Gson library to deal with Json fies
- Maven capabilities (json23plet is a Maven project)

Apache Jena

- Apache Jena is Java framework for building Semantic web
- The framework is composed of different APIs interacting together to process RDF data
- It is very complexed and requires much practice in order to understand how to work with
- We made the usage of creating RDF triples much simpler by wrap the Apache Jena with our triplet module

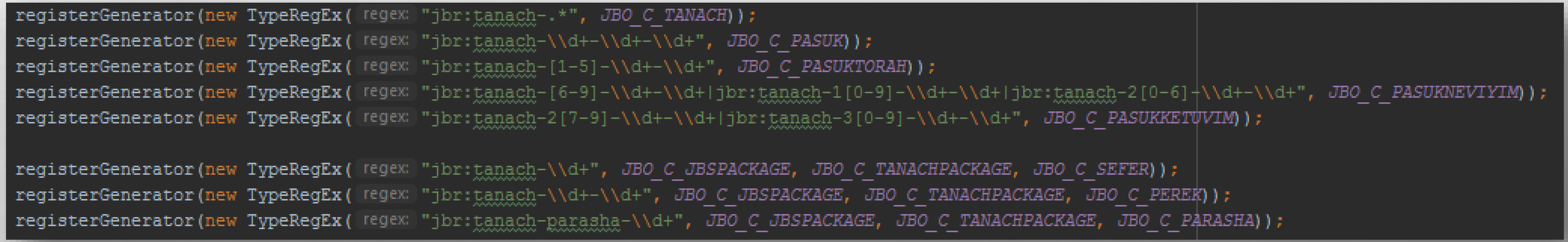
Apache Jena

RDF type - Regex generator

- While developing a Semantic Web domain there might be a lot of changes in Ontology, Predicates, Hierarchy, etc.
- We would like to reduce the impact of changes on the existing data and the development process
- Assume we decide to change the classes hierarchy, where should we change the flow?
  - In text2json – requires changes a lot of dedicated and different generators
  - Creating a generator for each class with json23plet will have the same problem
- Therefore, we developed the regex generator framework

Regex generator

- Contains “Rules”
- Rule will be activated on a Json object if they match, based on regular expressions matching
- Define all your frequently changed generators as rules in regex generator
- Now, if something has changed all is needed is to update the appropriate rule and rerun the regex generator
- Provides a simple and clean solution to the problem, as seen:



```
registerGenerator(new TypeRegEx( regex: "jbr:tanach-.*", JBO_C_TANACH));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-\\d+-\\d+-\\d+", JBO_C_PASUK));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-[1-5]-\\d+-\\d+", JBO_C_PASUKTORAH));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-[6-9]-\\d+-\\d+|jbr:tanach-1[0-9]-\\d+-\\d+|jbr:tanach-2[0-6]-\\d+-\\d+", JBO_C_PASUKNEVIYIM));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-2[7-9]-\\d+-\\d+|jbr:tanach-3[0-9]-\\d+-\\d+", JBO_C_PASUKKETUVIM));

registerGenerator(new TypeRegEx( regex: "jbr:tanach-\\d+", JBO_C_JBSPACKAGE, JBO_C_TANACHPACKAGE, JBO_C_SEFER));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-\\d+-\\d+", JBO_C_JBSPACKAGE, JBO_C_TANACHPACKAGE, JBO_C_PEREK));
registerGenerator(new TypeRegEx( regex: "jbr:tanach-parasha-\\d+", JBO_C_JBSPACKAGE, JBO_C_TANACHPACKAGE, JBO_C_PARASHA));
```

- **To get inside:**

  https://github.com/TechnionTDK/jbs-json23plet