Rule extraction from textual building regulations of the 
City of Vienna

Eszter Iklódi

TU Wien

Nov 23, 2021
Topics

- BRISE introduction
- Rule representation
- Data annotation process
- First results
- Next steps
- Discussion
Vienna: 165,000 building, 1M flats (dynamically expanding)
more than 200 tons of paper of archived building documents
150 building proposals / year / employee in MA37 (Baupolizei)
average processing time: 1 year
goal: complete digitalization from submission until approval
modern technologies: 3D-building modeling, AI, augmented reality
pioneer in e-governance, capital of digitalization
R & D project with 4.8M €
Building regulations

- Main building regulations (Bauordnung)
- Zoning plans (Flächenwidmungs- und Bebauungspläne)
- Textual regulations (Textliche Bestimmungen)
§75. (2) Die Gebäudehöhe hat, soweit sich nicht nach den Bestimmungen [...] des Bebauungsplanes eine andere Gebäudehöhe ergibt, zu betragen:
in Bauklasse I mindestens 2,5 m, höchstens 9 m,
in Bauklasse II mindestens 2,5 m, höchstens 12 m,
in Bauklasse III mindestens 9 m, höchstens 16 m,
in Bauklasse IV mindestens 12 m, höchstens 21 m,
in Bauklasse V mindestens 16 m, höchstens 26 m.
Zoning plans (Flächenwidmungs- und Bebauungspläne)

4.2. Auf den mit BB2 bezeichneten Grundflächen sind die zur Errichtung gelangenden Dächer entsprechend dem Stand der Technik als begrünte Flachdächer auszuführen.


4.6. Auf den mit BB7 bezeichneten Grundflächen ist in der Höhenlage der anschließenden Verkehrsfläche eine öffentliche Durchfahrt mit einer lichten Höhe von 3,0 m freizuhalten und zu dulden.
Main goal

To extract rules!
The rule extraction task

*Flachdächer bis zu einer Dachneigung von fünf Grad sind entsprechend dem Stand der technischen Wissenschaften zu begrünen.*

*Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.*

```json
{"modality": "obligation",
 "attributes": [
  {
    "type": "content",
    "name": "BegruenungDach",
    "value": null},

  {
    "type": "condition",
    "name": "Dachart",
    "value": "Flachdach"},

  {
    "type": "condition",
    "name": "DachneigungMax",
    "value": "5Grad"}]
}
```
Flachdächer bis zu einer Dachneigung von fünf Grad sind entsprechend dem Stand der technischen Wissenschaften zu begrünen.

Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.

{"modality": "obligation",
 "attributes": [
  {"type": "content",
   "name": "BegruenungDach",
   "value": null},

  {"type": "condition",
   "name": "Dachart",
   "value": "Flachdach"},

  {"type": "condition",
   "name": "DachneigungMax",
   "value": "5Grad"}]
}
Attribute names

- were extracted from an expert system
- originate from law
- annotated
  - 95 different attributes
  - 15,621 all together
- gold (10% of the docs)
  - 75 different attributes
  - 1,526 all together
- similar distribution
- unbalanced
- long-tail
Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.

{"modality": "obligation",
 "attributes": [
  {
   "type": "content",
   "name": "BegruenungDach",
   "value": "null"
  },

  {
   "type": "condition",
   "name": "Dachart",
   "value": "Flachdach"
  },

  {
   "type": "condition",
   "name": "DachneigungMax",
   "value": "5Grad"
  }]}
Attribute values

Different types

- **boolean**
  - ErrichtungGebaeude
  - BegruenungDach
  - AnordnungGaertnerischeAusgestaltung
  - VerbotWohnung
  - AufbautenZulaessig

- **text**
  - AusnahmeGaertnerischAuszugestaltende
  - EinfriedungLage
  - Planzeichen

- **number**
  - DurchfahrtBreite
  - DurchfahrtHoehe
  - GehsteigbreiteMin
  - AnteilDachbegruenung
Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.

{"modality": "obligation",
 "attributes": [
  {
   "type": "content",
   "name": "BegrüenungDach",
   "value": null,
  },
  {
   "type": "condition",
   "name": "Dachart",
   "value": "Flachdach"},
  {
   "type": "condition",
   "name": "DachneigungMax",
   "value": "5Grad"}]
}
The complete list:

- condition
  - *Flat roofs should be green roofs, unless they are glass roofs.*

- content
  - *Flat roofs should be green roofs, unless they are glass roofs.*

- conditionException
  - *Flat roofs should be green roofs, unless they are glass roofs.*

- contentException
  - *Windows are prohibited except for portholes.*
Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.

{"modality": "obligation", "attributes": [
    {"type": "content", "name": "BegrüenungDach", "value": null},
    {"type": "condition", "name": "Dachart", "value": "Flachdach"},
    {"type": "condition", "name": "DachneigungMax", "value": "5Grad"}]]}
Sentence modality - examples

- **obligation**
  - Flachdächer bis zu einer Dachneigung von fünf Grad sind entsprechend dem Stand der technischen Wissenschaften zu begrünen.
  - Flat roofs with a pitch not exceeding 5 degrees must be greened using state of the art technologies.

- **prohibition**
  - Die Herstellung von Einfriedungen ist untersagt.
  - The creation of an enclosure is prohibited.

- **permission**
  - Die Errichtung von Hochhäusern ist nicht zulässig.
  - The construction of skyscrapers are allowed.
Annotation process

- 2 phases
  - Label annotation (Sept)
  - Break (Oct)
  - Full annotation (Nov-Dec)
    - Modality & Type
    - Values - at review time

- Annotators
  - 6 university students
  - 140 hours each

- Materials
  - Guideline for all the attributes (47 pages)
  - 2 kick-off workshops
  - Some follow-up workshops
  - Slack-channel
Annotation process

- **Data characteristics**
  - around 1500 documents
  - 6-160 sentences / document

- **Sentence-based document distribution**
  - Multiway number partitioning: https://en.wikipedia.org/wiki/Multiway_number_partitioning
  - Karmarkar-karp algorithm: https://github.com/fuglede/numberpartitioning
Assignment requirements

Requirements

- 2 annotators / document
- batch-wise assignment
- 5 different batch types (everyone with everyone)
## Batch types

<table>
<thead>
<tr>
<th>cycle</th>
<th>annotator 1</th>
<th>annotator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>01</td>
<td>02</td>
</tr>
<tr>
<td>0</td>
<td>03</td>
<td>04</td>
</tr>
<tr>
<td>0</td>
<td>05</td>
<td>06</td>
</tr>
<tr>
<td>1</td>
<td>04</td>
<td>06</td>
</tr>
<tr>
<td>1</td>
<td>02</td>
<td>05</td>
</tr>
<tr>
<td>1</td>
<td>01</td>
<td>03</td>
</tr>
<tr>
<td>2</td>
<td>03</td>
<td>05</td>
</tr>
<tr>
<td>2</td>
<td>01</td>
<td>04</td>
</tr>
<tr>
<td>2</td>
<td>02</td>
<td>06</td>
</tr>
<tr>
<td>3</td>
<td>03</td>
<td>06</td>
</tr>
<tr>
<td>3</td>
<td>01</td>
<td>05</td>
</tr>
<tr>
<td>3</td>
<td>02</td>
<td>04</td>
</tr>
<tr>
<td>4</td>
<td>04</td>
<td>05</td>
</tr>
<tr>
<td>4</td>
<td>01</td>
<td>06</td>
</tr>
<tr>
<td>4</td>
<td>02</td>
<td>03</td>
</tr>
</tbody>
</table>
Pre-fill suggested attributes by rule-based system (high precision)
Doing parallel review & mark already gold sentences (with fuzzy match)
### Label review example

<table>
<thead>
<tr>
<th>Sentence_ID</th>
<th>Review</th>
<th>Sentence</th>
<th>Kategorie</th>
<th>Merkmal</th>
<th>Count</th>
<th>Given by</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>E141_B_D</td>
<td>Done</td>
<td>Bestimmungen ohne Beschränkung des Gebietebevors mit dem Flurzeichen 6B:</td>
<td>Lage_Gebiete, Planzeichen</td>
<td>Planzeichen</td>
<td>101</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td>E141_B_D</td>
<td>question</td>
<td>Der höchste Punkt der zur Erreichung gelangenden Dächer darf die ausgeführte Gebäudefläche um höchstens 4,5 m überragen.</td>
<td>Dach</td>
<td>AbschlussDachMaxlengGebäude</td>
<td>01</td>
<td>04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>202</td>
<td>01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>202</td>
<td>04</td>
</tr>
</tbody>
</table>
Assigned and annotated documents

Assigned documents
- 400 assigned documents
- 138-140 documents / annotator
- 3457-3460 sentences / annotator

Annotated documents
- avg. 117 annotated documents / annotator
- 300 docs annotated twice (goal achieved)
- 372 docs annotated once
Reviewed documents

- Currently 46 reviewed documents with 1247 sentences (= gold)
- Review time beginning: 2 min / sentence
- Review time later: 1 min / sentence
Gathered information

- `gen_attributes_on_annotation`: attributes suggested by the rule-based system
- `already_gold_on_annotation`: if gold already existed for the sentence at annotation time
- `annotated_attributes`: attributes annotated
Attribute distribution

Gold

Annotated
## Top 20 attribute frequency

### Gold

<table>
<thead>
<tr>
<th>attribute</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>318</td>
</tr>
<tr>
<td>Flächen</td>
<td>103</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>86</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>59</td>
</tr>
<tr>
<td>Dachart</td>
<td>56</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>54</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>51</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>45</td>
</tr>
<tr>
<td>BegrufnungDach</td>
<td>39</td>
</tr>
<tr>
<td>WidmungInMehrerenEbenen</td>
<td>38</td>
</tr>
</tbody>
</table>

### Annotated

<table>
<thead>
<tr>
<th>attribute</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>2881</td>
</tr>
<tr>
<td>Flächen</td>
<td>1070</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>947</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>575</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>553</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>523</td>
</tr>
<tr>
<td>GebaeudeHoeheMax</td>
<td>464</td>
</tr>
<tr>
<td>Dachart</td>
<td>434</td>
</tr>
<tr>
<td>GebaeudeHoeheArt</td>
<td>431</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>412</td>
</tr>
</tbody>
</table>

### Includes

- UnterbrechungGeschlosseneBauweise *
- StrassenbreiteMin
- ErrichtungGebaeude *
- AufbautenZulaessig
<table>
<thead>
<tr>
<th>Sentence_ID</th>
<th>Sentence</th>
<th>Modality</th>
<th>Kategorie</th>
<th>Merkmal</th>
<th>Value</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>B141_7_0</td>
<td>Entlang der Fluchttlinien sind Gehstege mit jeweils mindestens 2,0 m Breite herzustellen.</td>
<td>obligation</td>
<td>Lage_Gelaende_Planzeichen</td>
<td>AnFluchttlinie</td>
<td>True</td>
<td>condition</td>
</tr>
<tr>
<td>B141_8_0</td>
<td>Bestimmungen ohne Bezeichnung des Geltungsbereichs mit dem Planzeichen BB:</td>
<td>permission</td>
<td>Lage_Gelaende_Planzeichen</td>
<td>Planzeichen</td>
<td>BB</td>
<td>condition</td>
</tr>
<tr>
<td>B141_9_0</td>
<td>Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte Gebäudehöhe um höchstens 4,5 m überragen.</td>
<td>prohibition</td>
<td>Dach</td>
<td>AbschlussDachMaxBezugGebaeude</td>
<td>4,5 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Die zur Errichtung gelangenden Dächer von Gebäuden mit einer bebauten Fläche von mehr als 12 m² sind bis zu einer Dachneigung von 15° entsprechend dem Stand der Technik zu begrünen.</td>
<td></td>
<td>Dach</td>
<td>BegruenungDach</td>
<td>True</td>
<td>content</td>
</tr>
<tr>
<td>B141_10_0</td>
<td>Mit Ausnahme der als Gemischtes Baugebiet/Betriebsbaugebiet gewidmeten Flächen darf die mit Nebengebäuden bebaute Grundfläche höchstens 30 m² je Bauplatz betrugen.</td>
<td></td>
<td>Flaeche</td>
<td>Flaechen</td>
<td>höchstens 30 m²</td>
<td>content</td>
</tr>
<tr>
<td>Sentence_ID</td>
<td>Review</td>
<td>Sentence</td>
<td>01</td>
<td>02</td>
<td>Modality</td>
<td>Review</td>
</tr>
<tr>
<td>-------------</td>
<td>-------</td>
<td>----------</td>
<td>----</td>
<td>----</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>8543_7.o</td>
<td>Divo</td>
<td>Eingang der Fluchtdälen sind Gehstege mit jeweils mindestens 2,3 m Breite herzustellen.</td>
<td></td>
<td></td>
<td>probabilis</td>
<td>permisible</td>
</tr>
<tr>
<td>8543_8.o</td>
<td>Questis</td>
<td>Bestimmungen ohne Berücksichtigung des Gebäudebereichs mit dem Flächen BR:</td>
<td></td>
<td></td>
<td>probabilis</td>
<td>permisible</td>
</tr>
<tr>
<td>8543_9.o</td>
<td></td>
<td>Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte</td>
<td></td>
<td></td>
<td>probabilis</td>
<td>permisible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gebäudehöhe um höchstens 4,5 m überragen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8543_10.o</td>
<td></td>
<td>Die zur Errichtung gelangenden Dächer von Gebäuden mit einer behaarten Fläche von</td>
<td></td>
<td></td>
<td>probabilis</td>
<td>permisible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mehr als 12 m² sind bis zu einer Dachneigung von 15° entsprechend dem Stand der</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Technik zu begrünen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8543_11.o</td>
<td></td>
<td>Mit Ausnahme der als Gemächtes Baugebiet/Betriebsgebiet gewidmeten Flächen</td>
<td></td>
<td></td>
<td>probabilis</td>
<td>permisible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>darf die mit Nebengebäuden behaute Grundfläche höchstens 30 m² je Bauplatz</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>betragen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Full rule extraction task

Possible workflow:

- Find attributes in a sentence
- Determine modality for a sentence given the attributes
- Find values for the attributes
- Determine types for all the values of all attributes

Pipeline
- Preprocessing (pdf to tokens)
- UD parsing
- Concept graphs (4lang)
- Deontic rules

Attribute extraction
- Sample dataset: 10 documents with 193 sentences annotated for only attributes
- Attribute extraction: $p=0.93$, $r=0.51$

Rule extraction
- Toy dataset: 6 fully annotated sentences
- Perfect performance
First experiments with the annotated data

<table>
<thead>
<tr>
<th>set</th>
<th>annotated docs</th>
<th>gold docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>train (80%)</td>
<td>295</td>
<td>36</td>
</tr>
<tr>
<td>valid (10%)</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>test (10%)</td>
<td>39</td>
<td>5</td>
</tr>
</tbody>
</table>

Gold set is currently too small for experiments. We used gold when already there, otherwise the unreviewed annotations.
## Rule-Based for top 10 - only gold

### Slightly improved ASAIL rules

<table>
<thead>
<tr>
<th>attribute</th>
<th>precision</th>
<th>recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>0.98</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td>Flaechen</td>
<td>0.95</td>
<td>0.27</td>
<td>0.42</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>0.89</td>
<td>0.15</td>
<td>0.25</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>1.00</td>
<td>0.45</td>
<td>0.62</td>
</tr>
<tr>
<td>Dachart</td>
<td>0.98</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>0.91</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>0.55</td>
<td>0.77</td>
<td>0.64</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>1.00</td>
<td>0.65</td>
<td>0.79</td>
</tr>
<tr>
<td>BegruenungDach</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>WidmungInMehrerenEbenen</td>
<td>0.88</td>
<td>0.45</td>
<td>0.60</td>
</tr>
</tbody>
</table>

| macro-avg (w/o VerkehrsflaecheID)                  | 0.94      | 0.57   | 0.68 |
| micro-avg (w/o VerkehrsflaecheID)                  | 0.95      | 0.57   | 0.69 |
Rule-Based for top 10 - all annotated

Slightly improved ASAIL rules

<table>
<thead>
<tr>
<th>attribute</th>
<th>precision</th>
<th>recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>0.91</td>
<td>0.82</td>
<td>0.87</td>
</tr>
<tr>
<td>Flaechen</td>
<td>0.89</td>
<td>0.24</td>
<td>0.38</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>0.78</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>0.49</td>
<td>0.81</td>
<td>0.61</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>0.91</td>
<td>0.57</td>
<td>0.70</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>0.88</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>GebaeudeHoeheMax</td>
<td>0.94</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>Dachart</td>
<td>0.91</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>GebaeudeHoeheArt</td>
<td>0.92</td>
<td>0.44</td>
<td>0.59</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>0.95</td>
<td>0.70</td>
<td>0.80</td>
</tr>
</tbody>
</table>

| macro-avg (w/o VerkehrsflaecheID)             | 0.90      | 0.55   | 0.64 |
| micro-avg (w/o VerkehrsflaecheID)             | 0.89      | 0.58   | 0.67 |
Naive Bayes for top 3

Simple Naive Bayes implementation (from NLP lecture)
Best setup: filter stopword + stemming + balance train data

<table>
<thead>
<tr>
<th>attribute</th>
<th>precision</th>
<th>recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>.88</td>
<td>.71</td>
<td>.79</td>
</tr>
<tr>
<td>Flaechen</td>
<td>.25</td>
<td>.95</td>
<td>.40</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>.24</td>
<td>.95</td>
<td>.38</td>
</tr>
</tbody>
</table>
Best setup: filter stopword + stemming + min occur=5

<table>
<thead>
<tr>
<th>attribute</th>
<th>precision</th>
<th>recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>.94</td>
<td>.93</td>
<td>.94</td>
</tr>
<tr>
<td>Flaechen</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>.69</td>
<td>.78</td>
<td>.74</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>.86</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>.79</td>
<td>.90</td>
<td>.84</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>.93</td>
<td>.90</td>
<td>.92</td>
</tr>
<tr>
<td>GebaeudeHoeheMax</td>
<td>.77</td>
<td>.85</td>
<td>.81</td>
</tr>
<tr>
<td>Dachart</td>
<td>.88</td>
<td>.92</td>
<td>.90</td>
</tr>
<tr>
<td>GebaeudeHoeheArt</td>
<td>.95</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>.95</td>
<td>1.00</td>
<td>.98</td>
</tr>
</tbody>
</table>
# Compare label extraction methods

## F1 scores

<table>
<thead>
<tr>
<th>attribute</th>
<th>rule-based</th>
<th>BoW+LogReg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planzeichen</td>
<td>0.87</td>
<td>0.94</td>
</tr>
<tr>
<td>Flaechen</td>
<td>0.38</td>
<td>0.79</td>
</tr>
<tr>
<td>WidmungUndZweckbestimmung</td>
<td>0.23</td>
<td>0.74</td>
</tr>
<tr>
<td>VerkehrsflaecheID</td>
<td>0.61</td>
<td>0.90</td>
</tr>
<tr>
<td>AnFluchtlinie</td>
<td>0.70</td>
<td>0.84</td>
</tr>
<tr>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>GebaeudeHoeheMax</td>
<td>0.45</td>
<td>0.81</td>
</tr>
<tr>
<td>Dachart</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td>GebaeudeHoeheArt</td>
<td>0.59</td>
<td>0.95</td>
</tr>
<tr>
<td>VorkehrungBepflanzung</td>
<td>0.80</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Rule-Based VALUE extraction for top 10

Semi-manual analysis on gold attributes (1247 sentences)

<table>
<thead>
<tr>
<th>frequency</th>
<th>attribute</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>318</td>
<td>Planzeichen</td>
<td>p=1.0, r=0.9</td>
</tr>
<tr>
<td>103</td>
<td>Flaechen</td>
<td>p=1.0, r=0.9</td>
</tr>
<tr>
<td>86</td>
<td>WidmungUndZweckbestimmung</td>
<td>p=1.0, r=0.9</td>
</tr>
<tr>
<td>59</td>
<td>AnFluchtlinie</td>
<td>always True</td>
</tr>
<tr>
<td>56</td>
<td>Dachart</td>
<td>p=1.0, r=1.0</td>
</tr>
<tr>
<td>54</td>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>always True</td>
</tr>
<tr>
<td>51</td>
<td>VerkehrsflaecheID</td>
<td>API?</td>
</tr>
<tr>
<td>45</td>
<td>VorkehrungBepflanzung</td>
<td>p=1.0, r=0.9</td>
</tr>
<tr>
<td>39</td>
<td>BegruenungDach</td>
<td>always True</td>
</tr>
<tr>
<td>38</td>
<td>WidmungInMehrerenEbenen</td>
<td>p=1.0, r=1.0</td>
</tr>
</tbody>
</table>

With regex (at most 1.5 hours spent / attribute)
Attribute value is encoded by itself
Semi-manual analysis on gold attributes (1247 sentences)

<table>
<thead>
<tr>
<th>frequency</th>
<th>attribute</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>318</td>
<td>Planzeichen</td>
<td>always condition</td>
</tr>
<tr>
<td>103</td>
<td>Flächen</td>
<td>TBD</td>
</tr>
<tr>
<td>86</td>
<td>WidmungUndZweckbestimmung</td>
<td>TBD</td>
</tr>
<tr>
<td>59</td>
<td>AnFluchtlinie</td>
<td>always condition</td>
</tr>
<tr>
<td>56</td>
<td>Dachart</td>
<td>p=1.0, r=0.95</td>
</tr>
<tr>
<td>54</td>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>p=1.0, r=0.95</td>
</tr>
<tr>
<td>51</td>
<td>VerkehrsflaecheID</td>
<td>always condition</td>
</tr>
<tr>
<td>45</td>
<td>VorkehrungBepflanzung</td>
<td>always content</td>
</tr>
<tr>
<td>39</td>
<td>BegruenungDach</td>
<td>always content</td>
</tr>
<tr>
<td>38</td>
<td>WidmungInMehrerenEbenen</td>
<td>always content</td>
</tr>
</tbody>
</table>

With regex (at most 1.5 hours spent / attribute)
Attribute type is encoded by itself
Rule-Based MODALITY extraction for top 10

Manual analysis on gold sentences (1247 sentences)

<table>
<thead>
<tr>
<th>frequency</th>
<th>attribute</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>318</td>
<td>Planzeichen</td>
<td>everything</td>
</tr>
<tr>
<td>103</td>
<td>Flaechen</td>
<td>always obligation</td>
</tr>
<tr>
<td>86</td>
<td>WidmungUndZweckbestimmung</td>
<td>mostly obligation</td>
</tr>
<tr>
<td>59</td>
<td>AnFluchtlinie</td>
<td>mostly obligation</td>
</tr>
<tr>
<td>56</td>
<td>Dachart</td>
<td>always obligation</td>
</tr>
<tr>
<td>54</td>
<td>AnordnungGaertnerischeAusgestaltung</td>
<td>always obligation</td>
</tr>
<tr>
<td>51</td>
<td>VerkehrsflaecheID</td>
<td>always obligation</td>
</tr>
<tr>
<td>45</td>
<td>VorkehrungBepflanzung</td>
<td>always obligation</td>
</tr>
<tr>
<td>39</td>
<td>BegruenungDach</td>
<td>always obligation</td>
</tr>
<tr>
<td>38</td>
<td>WidmungInMehrerenEbenen</td>
<td>always obligation</td>
</tr>
</tbody>
</table>

Attribute encodes sentences modality by itself
Next steps

- Try Deep Learning baselines (BERT)
- Generate rules with POTATO
- Try learning rules (e.g. in DNF with Integer Programming and Column Generation)
Any ideas?
Thank you for you attention :}

---

Eszter Iklódi

Rule extraction from textual building regulations of the City of Vienna

Nov 23, 2021 44 / 44