

Rule extraction from textual building regulations of the City of Vienna

Eszter Iklódi

TU Wien

Nov 23, 2021

- BRISE introduction
- Rule representation
- Data annotation process
- First results
- Next steps
- Discussion

The BRISE project - Motivation

- 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 €

Video: <https://vimeo.com/600019926?dnt=1>

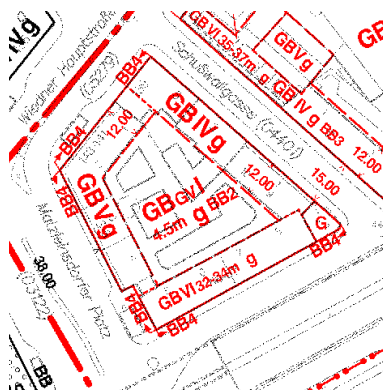
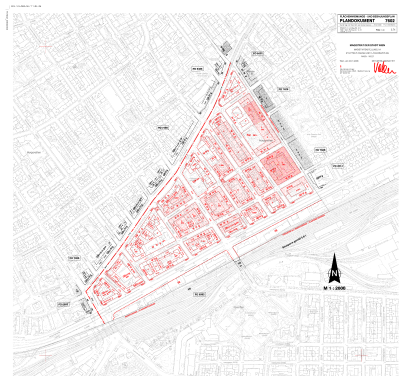
- Main building regulations (Bauordnung)
- Zoning plans (Flächenwidmungs- und Bebauungspläne)
- Textual regulations (Textliche Bestimmungen)

Main building regulations (Bauordnung)

§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)



Textual regulations (Textliche Bestimmungen)

4. 1. Auf den mit **BB1** bezeichneten Grundflächen ist die Errichtung von unterirdischen und oberirdischen Gebäuden untersagt.
4. 2. Auf den mit **BB2** bezeichneten Grundflächen sind die zur Errichtung gelangenden Dächer entsprechend dem Stand der Technik als begrünte Flachdächer auszuführen.
4. 3. Auf den mit **BB3** bezeichneten Grundflächen ist die Errichtung von Staffelgeschossen an den Straßenfronten zulässig.
4. 4. Entlang der mit **BB4** bezeichneten Baulinien dürfen keine Hauptfenster von Aufenthaltsräumen von Wohnungen im Erdgeschoss zu den Verkehrsflächen hin orientiert werden.
4. 5. Auf den mit **BB5** bezeichneten und als Grünland/Erholungsgebiet/Sport- und Spielplätze gewidmeten Grundflächen dürfen Gebäude bis zu einer Gebäudehöhe von 4,5 m errichtet werden. Das Ausmaß der bebauten Fläche darf maximal 10 v.H. der gesamten Grundfläche betragen.
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.

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

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

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

Attribute names

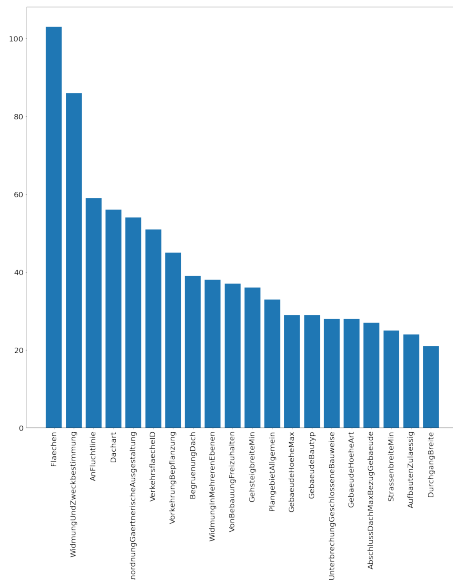
Flachdächer bis zu einer Dachneigung von fünf Grad sind entsprechend dem Stand der technischen Wissenschaften zu begrünen.

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

```
{"modality": "obligation",  
  "attributes": [  
    {"type": "content",  
      "name": "BegrueungDach",  
      "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
 - ▶ 1526 all together
- similar distribution
- unbalanced
- long-tail



Attribute values

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"}]}
```

Different types

- boolean

- ▶ ErrichtungGebaeude
- ▶ BegrueungDach
- ▶ AnordnungGaertnerischeAusgestaltung
- ▶ VerbotWohnung
- ▶ AufbautenZulaessig

- text

- ▶ AusnahmeGaertnerischAuszugestaltende
- ▶ EinfriedungLage
- ▶ Planzeichen

- number

- ▶ DurchfahrtBreite
- ▶ DurchfahrtHoehe
- ▶ GehsteigbreiteMin
- ▶ AnteilDachbegrueung

Attribute types

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"}]}
```

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.*

Sentence modality

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"}]}
```


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

- 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

cycle	annotator 1	annotator 2
0	01	02
0	03	04
0	05	06
1	04	06
1	02	05
1	01	03
2	03	05
2	01	04
2	02	06
3	03	06
3	01	05
3	02	04
4	04	05
4	01	06
4	02	03

Label annotation example

Sentence_ID	Sentence	Class1		Class2	
		Kategorie	Merkmal	Kategorie	Merkmal
8250_16_0	In Verbindung der mit a-b und c-d bezeichneten Buchstabenpaare wird in der mit BB11 (Punkt 4.11.) definierten Höhenlage ein Durchgang mit einer Breite von 8,0 m angeordnet.	Lage_Gelaende_Planzeichen	Planzeichen	Laubengaenge_Durchfahrten_Arkaden	DurchgangBreite

- Pre-fill suggested attributes by rule-based system (high precision)
- Doing parallel review & mark already gold sentences (with fuzzy match)

ID	Sentence	Class1	
		Kategorie	Merkmal
8141_8_0	Bestimmungen ohne Bezeichnung des Geltungsbereichs mit dem Planzeichen BB:	Lage_Gelaende_Planzeichen	Planzeichen
8141_9_0	Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte Gebäudehöhe um höchstens 4,5 m überragen.	Dach	AbschlussDachMaxBezugGebaeude

Label review example

Sentence_ID	Review	Sentence	Kategorie	Merkmal	Count	Given by	Review	Kategorie	Merkmal	Count	Given by	Review
8141_8_0	Done	Bestimmungen ohne Bezeichnung des Geltungsbereichs mit den Planzeichen BB:	Lage_Gebäude_Planzeichen	Planzeichen	1 01	OK	OK	Lage_Gebäude_Planzeichen	Planzeichen	1 01	OK	OK
8141_9_0	Done	Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte Gebäudehöhe um höchstens 4,5 m überragen.	Dach	AbschlussDachMaxBrenzGebäude	01	OK	OK	Höhe	GebäudeHöheArt	01	OK	OK
					2 02	OK	OK			2 02	OK	OK

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

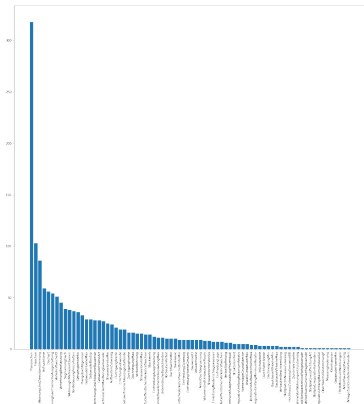
- Currently 46 reviewed documents with 1247 sentences (= gold)
- Review time beginning: 2 min / sentence
- Review time later: 1 min / sentence

Gathered information

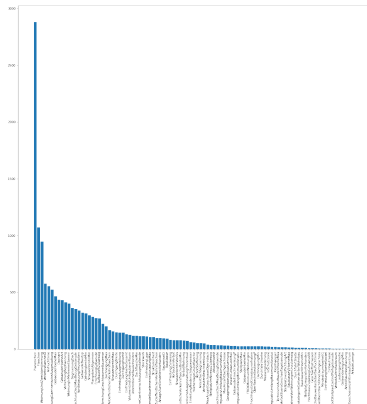
- `gen_attributes_on_annotation`: attributes suggested by the rule-based system
- `already_gold_on_annotation`: if gold already existed for the sentence at annotation time
- `annotated_attributes`: attributes annotated

Attribute distribution

Gold



Annotated



Top 20 attribute frequency

Gold

attribute	freq
Planzeichen	318
Flaechen	103
WidmungUndZweckbestimmung	86
AnFluchtlinie	59
Dachart	56
AnordnungGaertnerischeAusgestaltung	54
VerkehrsflaechelD	51
VorkehrungBepflanzung	45
BegruenungDach	39
WidmungInMehrerenEbenen	38
VonBebauungFreizuhalten	37
GehsteigbreiteMin	36
PlangebietAllgemein	33
GebaeudeHoeheMax	29
GebaeudeBautyp	29
<i>UnterbrechungGeschlosseneBauweise *</i>	28
GebaeudeHoeheArt	28
AbschlussDachMaxBezugGebaeude	27
StrassenbreiteMin	25
AufbautenZulaessig	24

Annotated

attribute	freq
Planzeichen	2881
Flaechen	1070
WidmungUndZweckbestimmung	947
VerkehrsflaechelD	575
AnFluchtlinie	553
AnordnungGaertnerischeAusgestaltung	523
GebaeudeHoeheMax	464
Dachart	434
GebaeudeHoeheArt	431
VorkehrungBepflanzung	412
WidmungInMehrerenEbenen	400
BegruenungDach	363
AbschlussDachMaxBezugGebaeude	352
VonBebauungFreizuhalten	340
GebaeudeBautyp	320
GehsteigbreiteMin	314
StrassenbreiteMin	298
PlangebietAllgemein	283
<i>ErrichtungGebaeude *</i>	274
AufbautenZulaessig	270

Full annotation example

Sentence_ID	Sentence	Modality	Kategorie	Merkmal	Value	Type
B141_7_0	Entlang der Fluchtlinien sind Gehsteige mit jeweils mindestens 2,0 m Breite herzustellen.	obligation	Lage_Gelaende_Planzeichen	AnFluchtlinie	True	condition
B141_8_0	Bestimmungen ohne Bezeichnung des Geltungsbereichs mit dem Planzeichen BB:	permission	Lage_Gelaende_Planzeichen	Planzeichen	BB	condition
B141_9_0	Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte Gebäudehöhe um höchstens 4,5 m überragen.	prohibition	Dach	AbschlussDachMaxBezugGebaeude	4,5 m	
B141_10_0	Die zur Errichtung gelangenden Dächer von Gebäuden mit einer bebauten Fläche von mehr als 12 m ² sind bis zu einer Dachneigung von 15° entsprechend dem Stand der Technik zu begrünen.		Dach	BegruenungDach	True	content
B141_11_0	Mit Ausnahme der als Gemischtes Baugebiet/Betriebsbaugebiet gewidmeten Flächen darf die mit Nebengebäuden bebaute Grundfläche höchstens 30 m ² je Bauplatz betragen.		Flaeche	Flaechen	höchstens 30 m ²	content

Full review example

Sentence_ID	Review	Sentence	01	02	Modality Review	Kategorie - 1	Merkmal - 1	Value - 1	01	02	Type Review
8141_7_0	Done	Entlang der Fluchtlinien sind Gehsteige mit jeweils mindestens 2,0 m Breite herzustellen.	prohibitives	permissives		Lage_Gelände_Planzeichen	AnFluchtlinie	True	condition	condition	condition
8141_8_0	Question	Bestimmungen ohne Bezeichnung des Geltungsbereichs mit dem Planzeichen BB:	prohibitives	permissives		Lage_Gelände_Planzeichen	Planzeichen	BB	condition	condition	condition
8141_9_0		Der höchste Punkt der zur Errichtung gelangenden Dächer darf die ausgeführte Gebäudehöhe um höchstens 4,5 m überragen.	prohibitives	permissives		Dach	AbschlussDachMaßBezugGebäude	4,5 m	EMPTY	EMPTY	
8141_10_0		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.	prohibitives	permissives		Dach	BegrünungDach	True	content	content	content
8141_11_0		Mit Ausnahme der als Genießerbauehler/Betriebsbauehler gewidmeten Flächen darf die mit Nebengebäuden bebaute Grundfläche höchstens 30 m² je Bauplatz betragen.	prohibitives	permissives		Fläche	Flächen	höchstens 30 m²	content	content	content

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

Proof of Concept - ASAIL paper

Recski, G., Lellmann, B., Kovacs, A. and Hanbury, A., 2021. Explainable Rule Extraction via Semantic Graphs.

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

set	annotated docs	gold docs
train (80%)	295	36
valid (10%)	37	5
test (10%)	39	5

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

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

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

attribute	precision	recall	F1
Planzeichen	.88	.71	.79
Flaechen	.25	.95	.40
WidmungUndZweckbestimmung	.24	.95	.38

BoW + Logistic Regression for top 10

Best setup: filter stopword + stemming + min occur=5

attribute	precision	recall	F1
Planzeichen	.94	.93	.94
Flaechen	.79	.79	.79
WidmungUndZweckbestimmung	.69	.78	.74
VerkehrsflaechelD	.86	.95	.90
AnFluchtlinie	.79	.90	.84
AnordnungGaertnerischeAusgestaltung	.93	.90	.92
GebaeudeHoeheMax	.77	.85	.81
Dachart	.88	.92	.90
GebaeudeHoeheArt	.95	.95	.95
VorkehrungBepflanzung	.95	1.00	.98

Compare label extraction methods

F1 scores

attribute	rule-based	BoW+LogReg
Planzeichen	0.87	0.94
Flaechen	0.38	0.79
WidmungUndZweckbestimmung	0.23	0.74
VerkehrsflaechelD	0.61	0.90
AnFluchtlinie	0.70	0.84
AnordnungGaertnerischeAusgestaltung	0.89	0.92
GebaeudeHoeheMax	0.45	0.81
Dachart	0.87	0.90
GebaeudeHoeheArt	0.59	0.95
VorkehrungBepflanzung	0.80	0.98

Rule-Based VALUE extraction for top 10

Semi-manual analysis on gold attributes (1247 sentences)

frequency	attribute	result
318	Planzeichen	p=1.0, r=0.9
103	Flaechen	p=1.0, r=0.9
86	WidmungUndZweckbestimmung	p=1.0, r=0.9
59	AnFluchtlinie	always True
56	Dachart	p=1.0, r=1.0
54	AnordnungGaertnerischeAusgestaltung	always True
51	VerkehrsflaecheID	API?
45	VorkehrungBepflanzung	p=1.0, r=0.9
39	BegruenungDach	always True
38	WidmungInMehrerenEbenen	p=1.0, r=1.0

With regex (at most 1.5 hours spent / attribute)
Attribute value is encoded by itself

Rule-Based TYPE extraction for top 10

Semi-manual analysis on gold attributes (1247 sentences)

frequency	attribute	result
318	Planzeichen	always condition
103	Flaechen	TBD
86	WidmungUndZweckbestimmung	TBD
59	AnFluchtlinie	always condition
56	Dachart	p=1.0, r=0.95
54	AnordnungGaertnerischeAusgestaltung	p=1.0, r=0.95
51	VerkehrsflaechelD	always condition
45	VorkehrungBepflanzung	always content
39	BegrueungDach	always content
38	WidmungInMehrerenEbenen	always content

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)

frequency	attribute	result
318	Planzeichen	everything
103	Flaechen	always obligation
86	WidmungUndZweckbestimmung	mostly obligation
59	AnFluchtlinie	mostly obligation
56	Dachart	always obligation
54	AnordnungGaertnerischeAusgestaltung	always obligation
51	VerkehrsflaecheID	everything
45	VorkehrungBepflanzung	always obligation
39	BegruenungDach	always obligation
38	WidmungInMehrerenEbenen	always obligation

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 :)