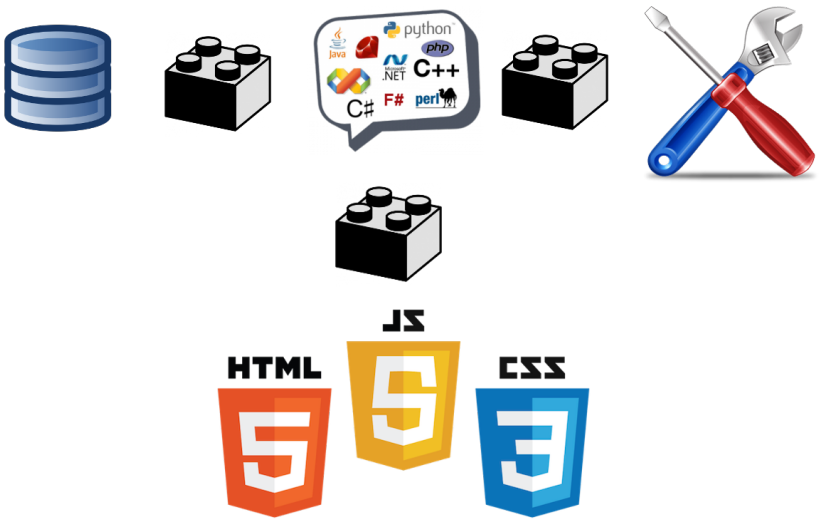
Simple Rule Inference from Configuration File

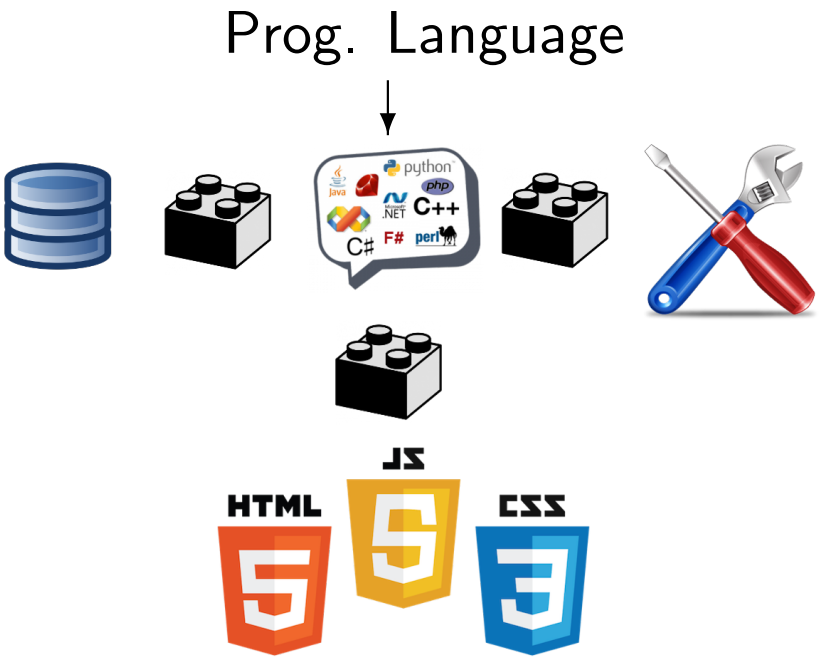
Kirill Levitin

July 21, 2015

How to build an enterprise software



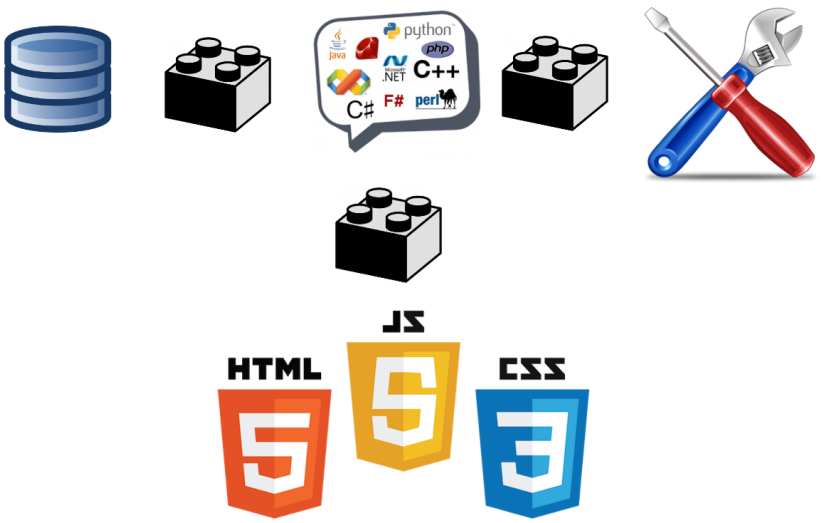
How to build an enterprise software



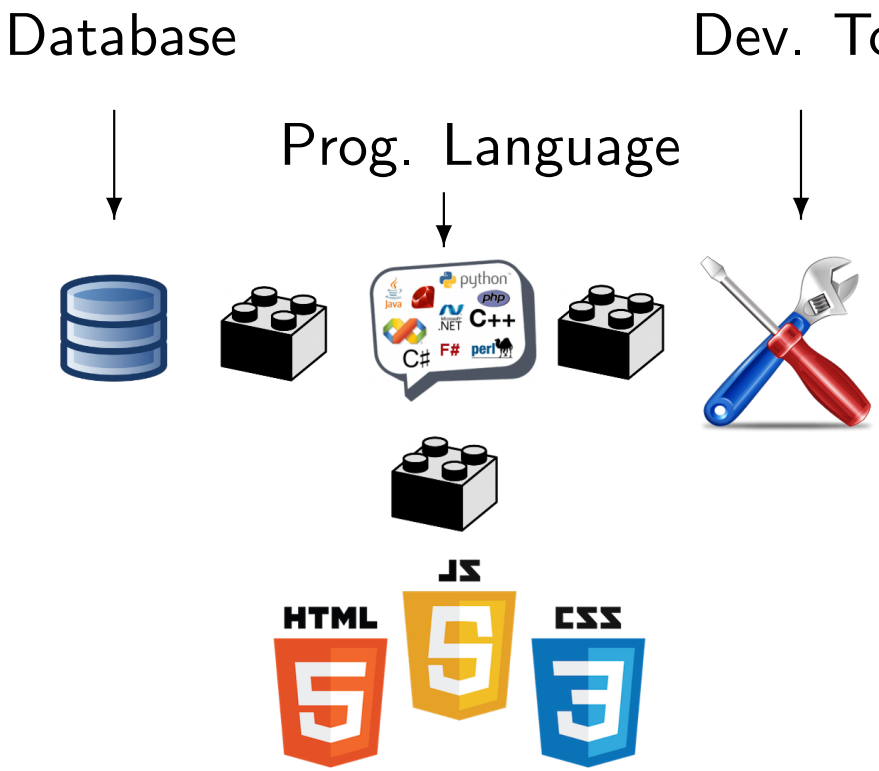
How to build an enterprise software

Database

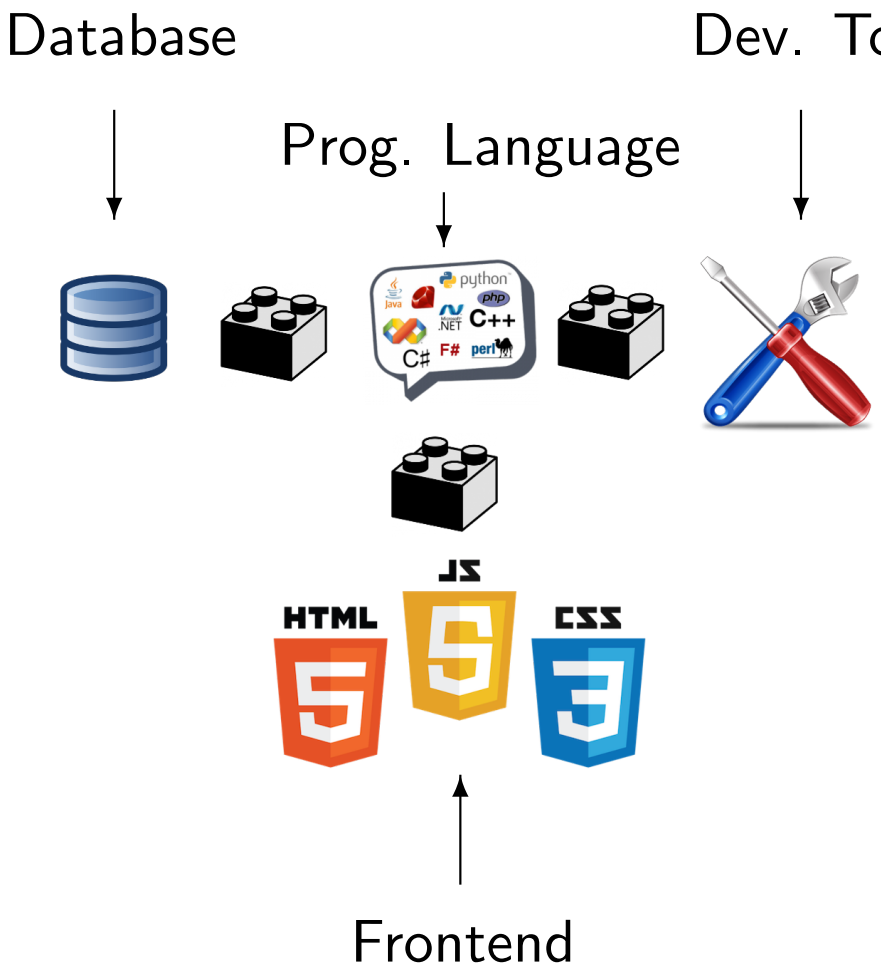
Prog. Language



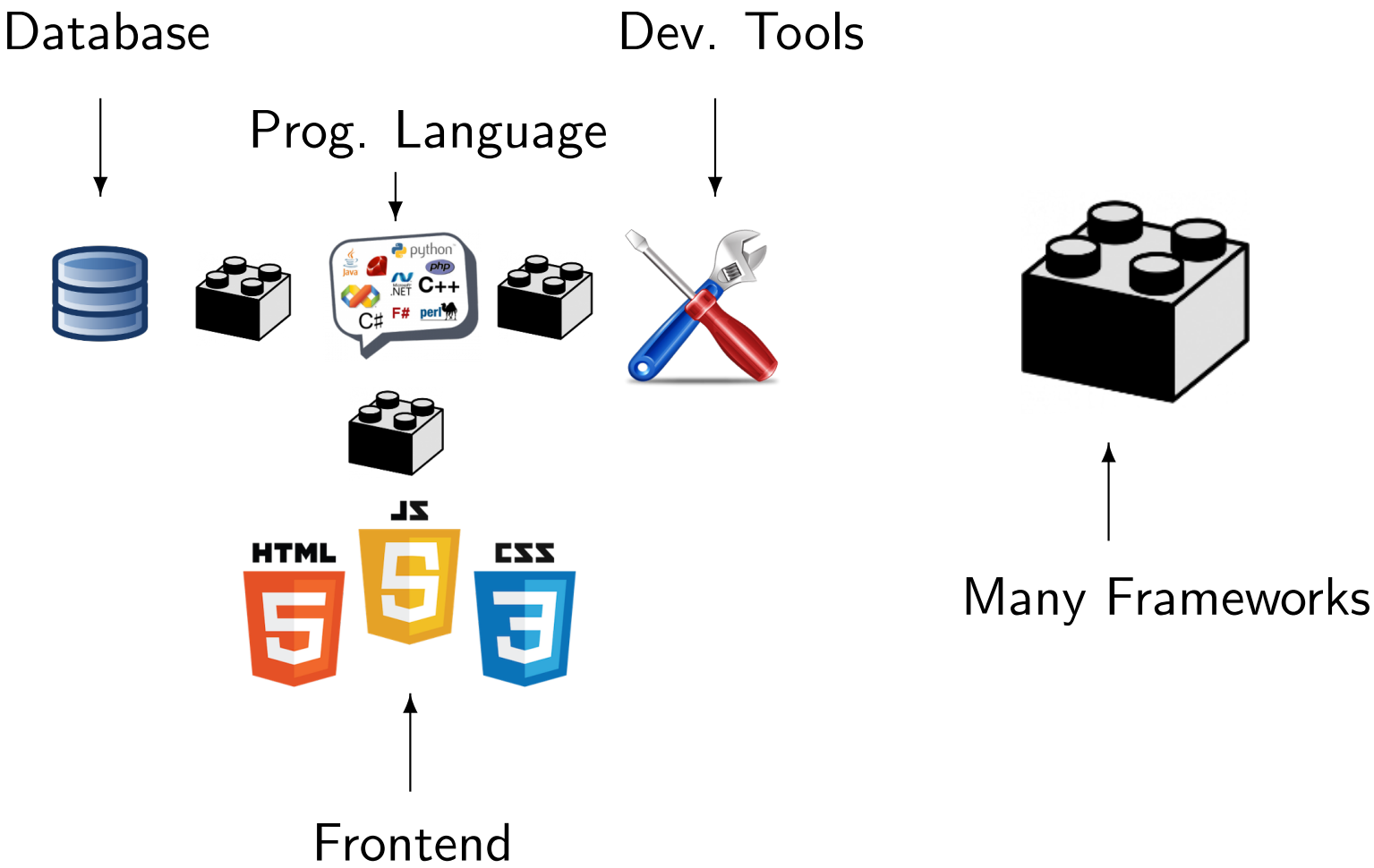
How to build an enterprise software



How to build an enterprise software



How to build an enterprise software



Small Example from J2EE Realization

Small Example from J2EE Realization

Many Frameworks

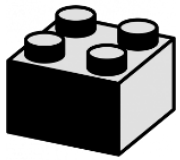
Software is not only a source code

Software is not only a source code

Programming Language

Software is not only a source code

Programming Language

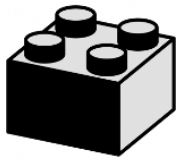


Frameworks

Software is not only a source code

Programming Language

XML Configuration Files



Frameworks

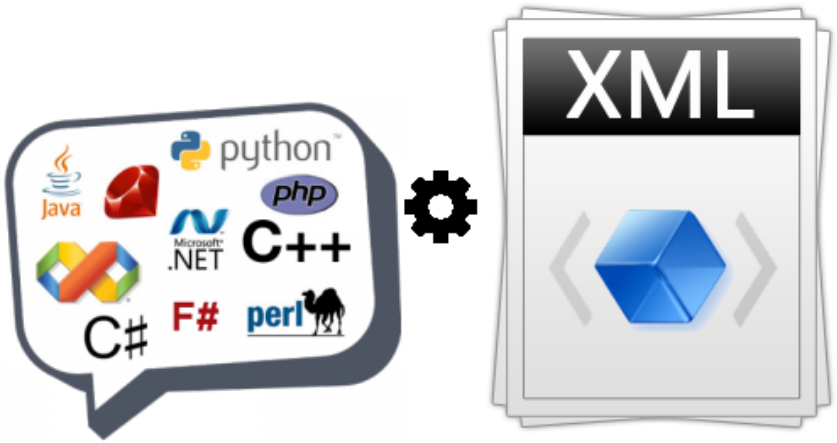
Motivation

Motivation

A **lot** of work has been done to analyse source code

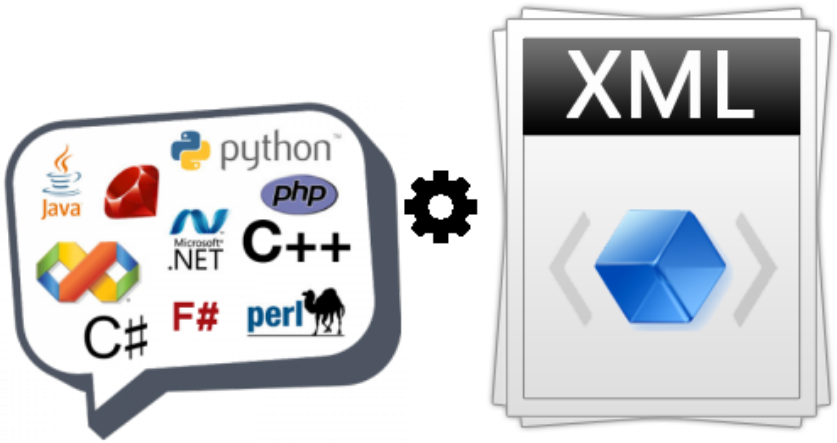
Little work has been dedicated to XML configuration files

Motivation



We will

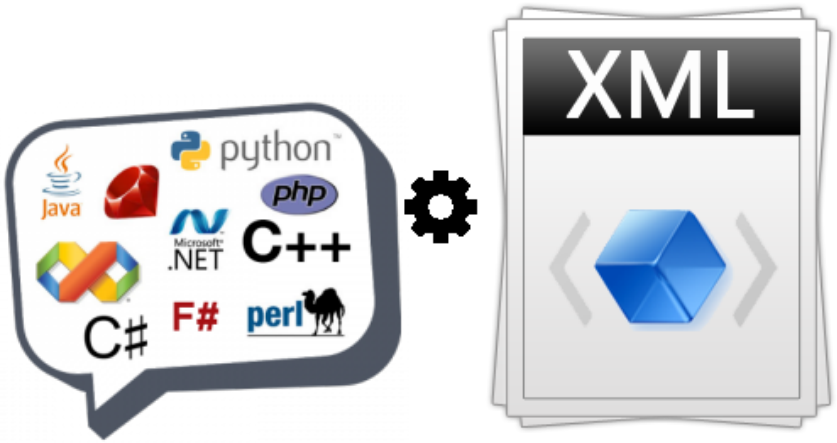
Motivation



We will

- analyse **both**, source code and XML files,

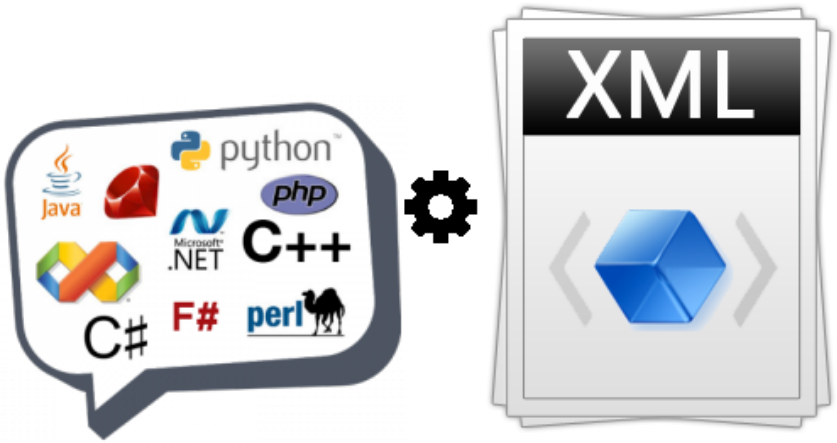
Motivation



We will

- analyse **both**, source code and XML files,
- **link** them together in one model,

Motivation



We will

- analyse **both**, source code and XML files,
- **link** them together in one model,
- discover **patterns**

Motivation

With discovered pattern we will **build tools** to help the developer in his everyday activities

Motivation

With discovered pattern we will **build tools** to help the developer in his everyday activities

- auto completion,

Motivation

With discovered pattern we will **build tools** to help the developer in his everyday activities

- auto completion,
- bug detection,

Motivation

With discovered pattern we will **build tools** to help the developer in his everyday activities

- auto completion,
- bug detection,
- ...

Example

```
<jpa:repositories base-package="ch.unibe.scg.persistence.repository" />
```

```
package ch.unibe.scg.persistence.repository;

public interface UserRepository extends JpaRepository<StudentEntity, Long> {

}
```

Example

```
<jpa:repositories base-package="ch.unibe.scg.persistence.repository" />
```

```
package ch.unibe.scg.persistence.repository;

public interface UserRepository extends JpaRepository<StudentEntity, Long> {

}
```

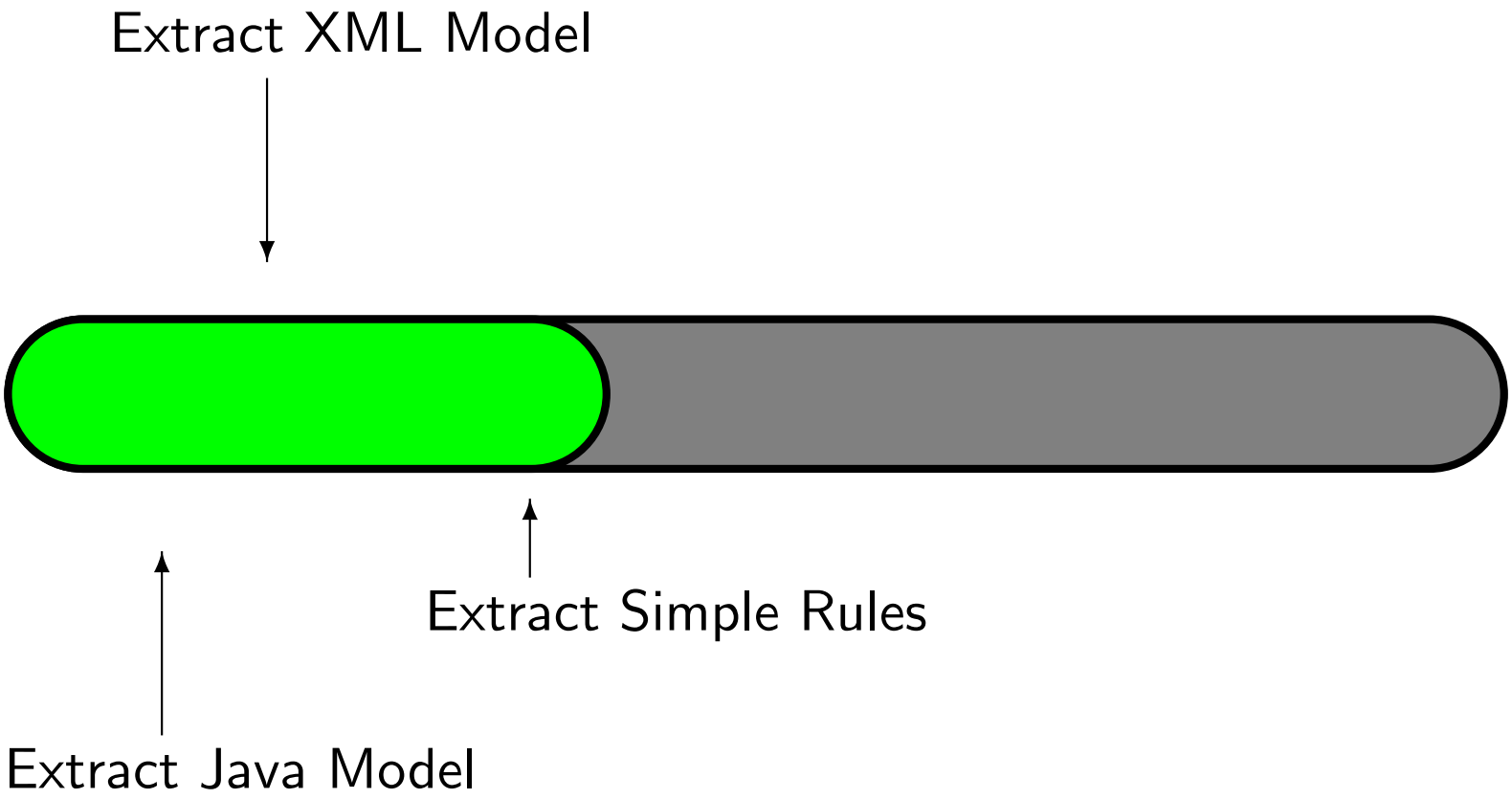
Example

```
<bean id="oracleDataSource" class="org.springframework....">
    ....
</bean>

<bean id="oracleEntityManagerFactory" class="org.springframework....">
    <property name="dataSource" ref="oracleDataSource" />
    ....
</bean>

<bean id="transactionManager" class="org.springframework....">
    <property name="entityManagerFactory" ref="oracleEntityManagerFactory" />
</bean>
```

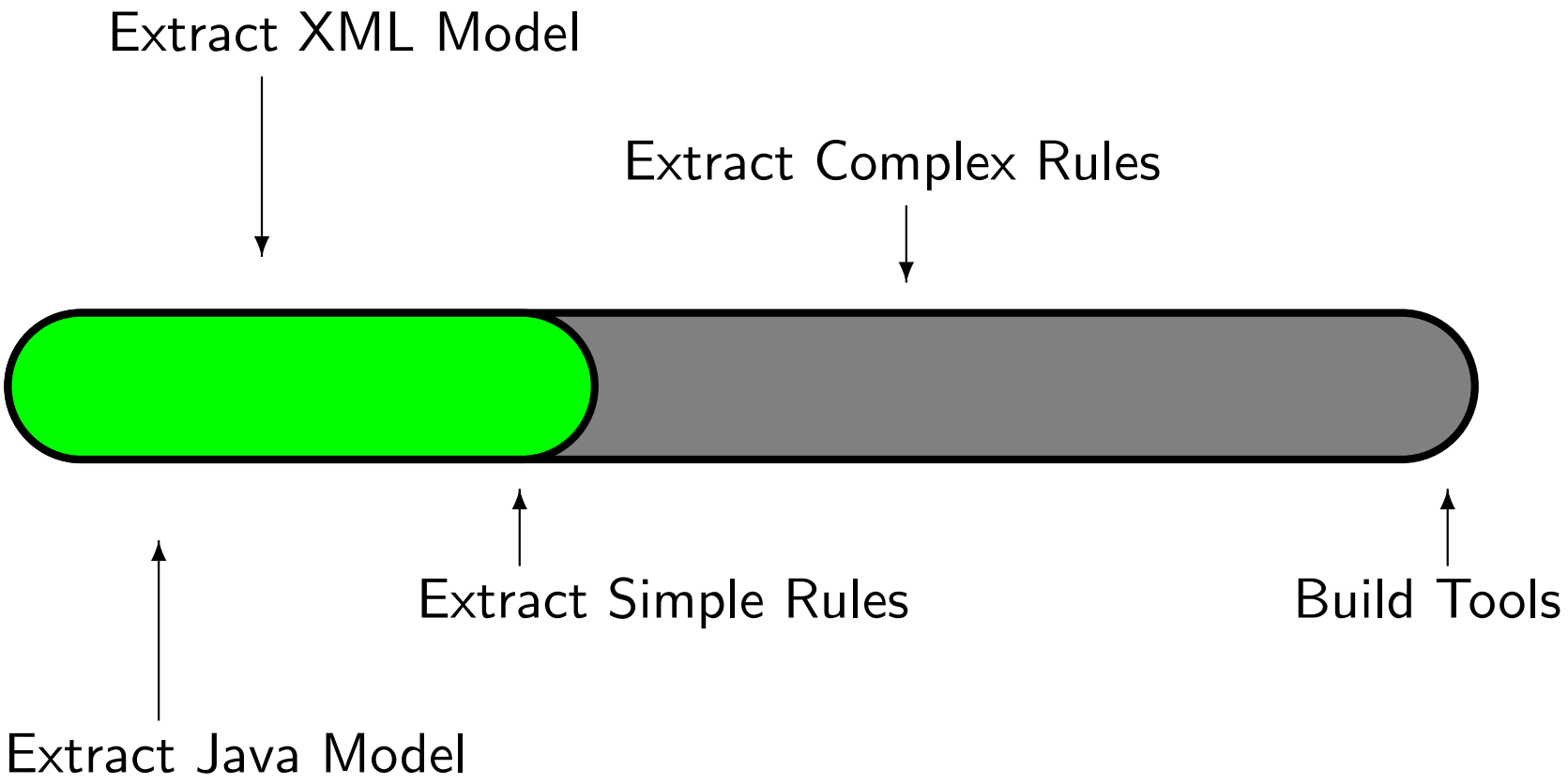
Example

```
<bean id="oracleDataSource" class="org.springframework....">
    .....
</bean>

<bean id="oracleEntityManagerFactory" class="org.springframework....">
    <property name="dataSource" ref="oracleDataSource" />
    .....
</bean>

<bean id="transactionManager" class="org.springframework....">
    <property name="entityManagerFactory" ref="oracleEntityManagerFactory" />
</bean>
```

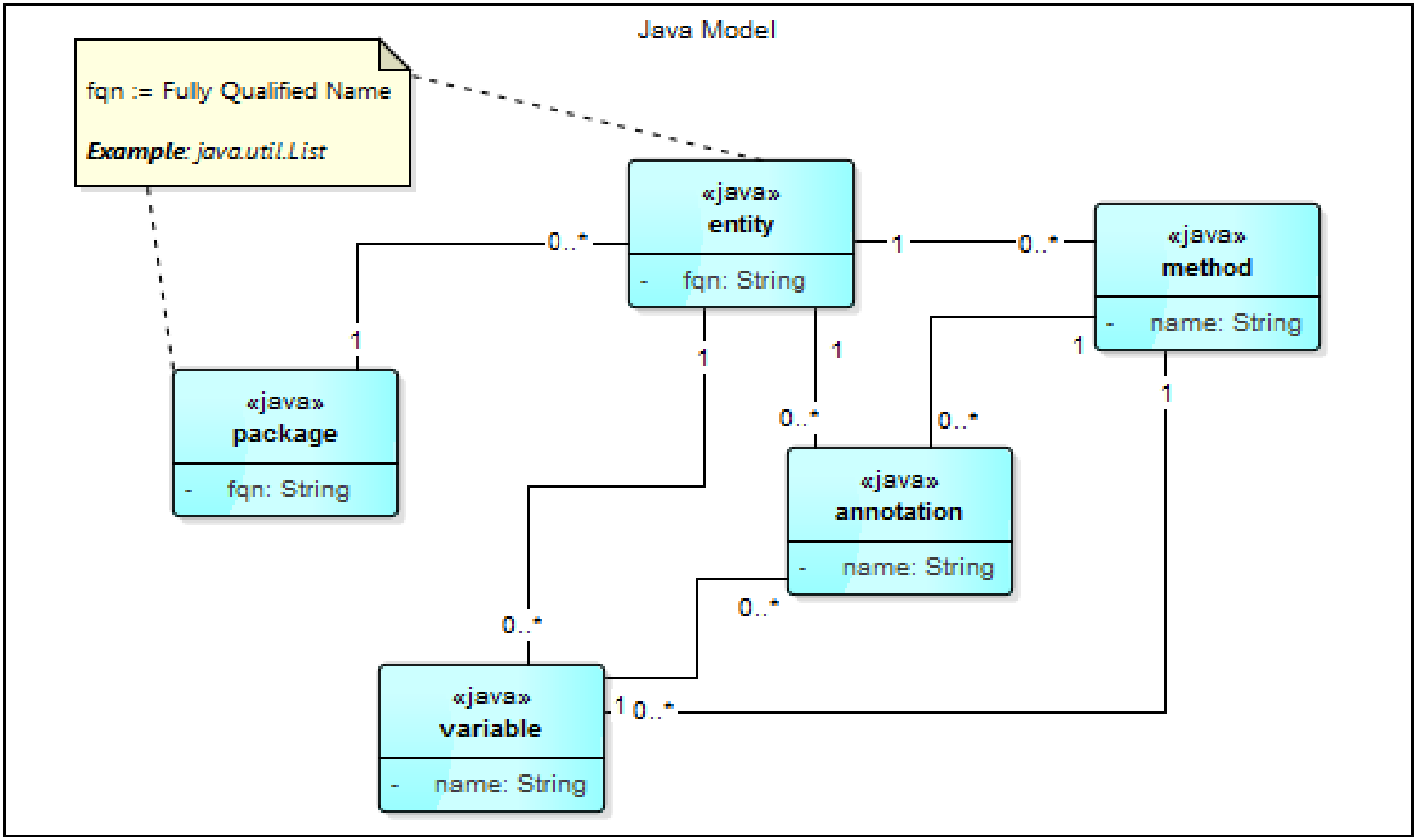
Progress



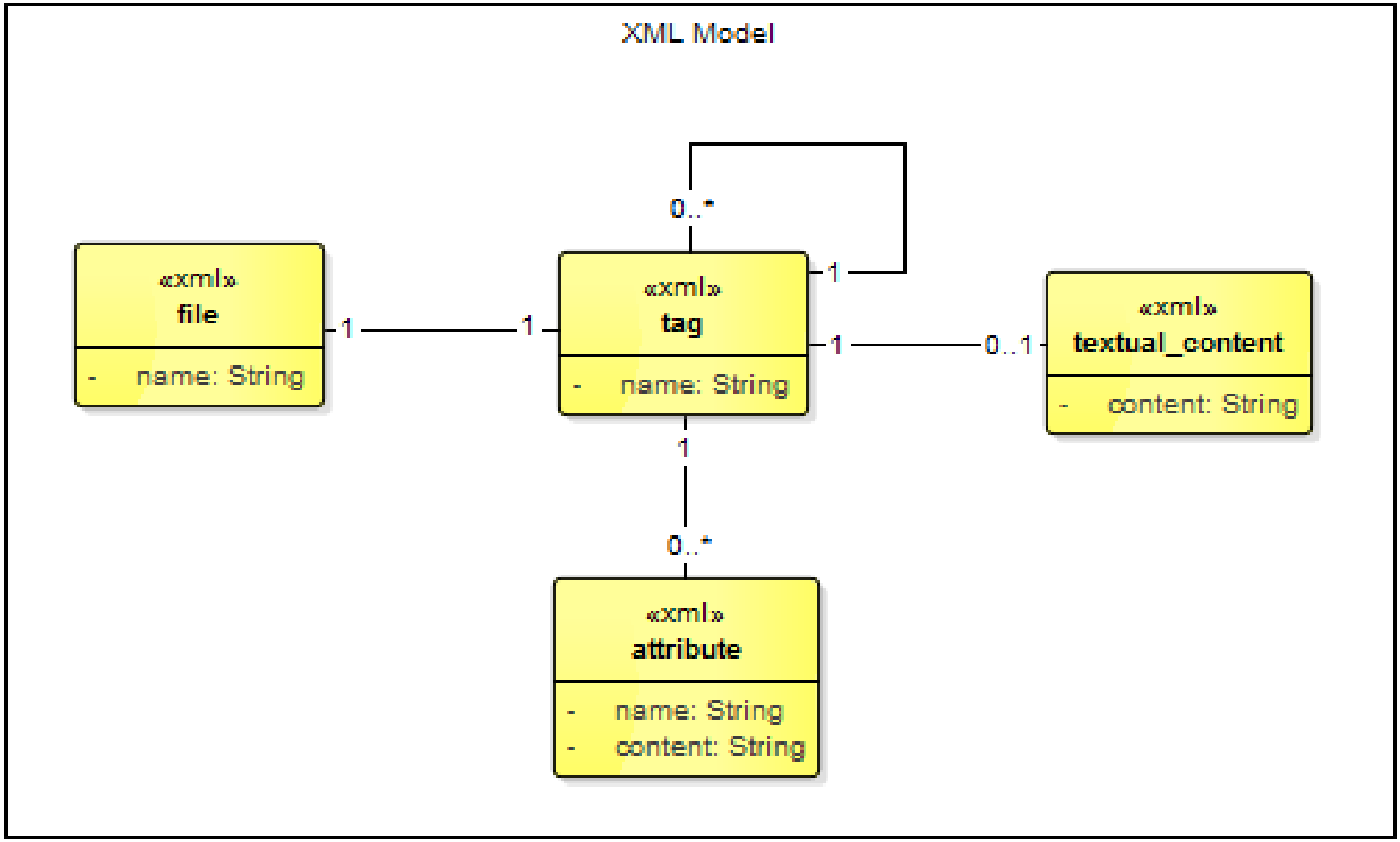
Progress


Extract XML Model
Extract Complex Rules
Extract Simple Rules
Build Tools
Extract Java Model

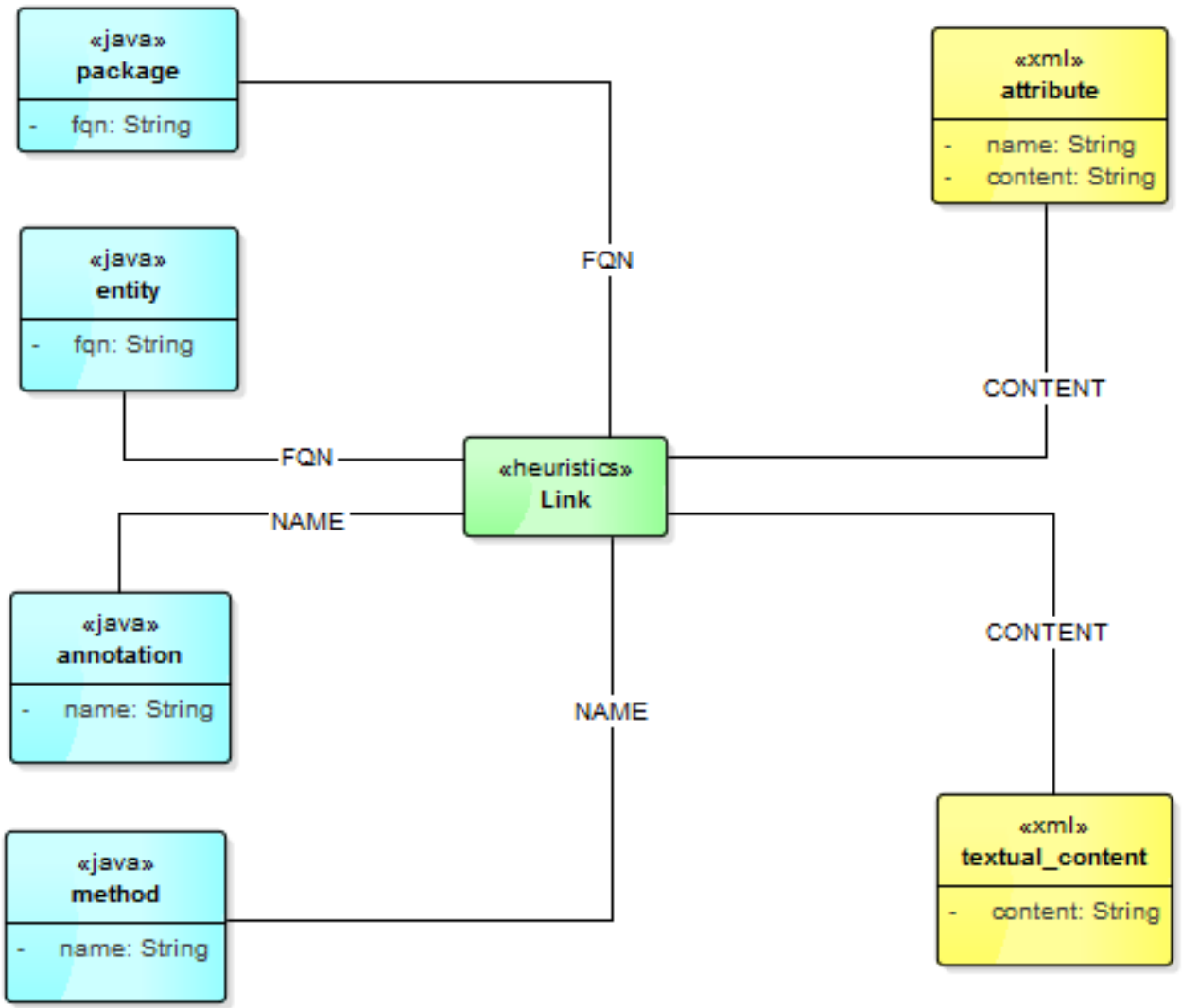
Extract Java Model



Extract XML Model



Link Java and XML Model



Calculate Support

$SUP\,(Link)$:= How often is the given Link's **template** found in all projects

Calculate Support

$SUP\,(Link)$:= How often is the given Link's **template** found in all projects

Example:
We have found $beans.bean.class = 4000$ in our 110 projects $\Rightarrow$

$$SUP\,(beans.bean.class) = 4000$$

Calculate Hit

$HIT\,(Link) :=$ How often is the given Link found in all projects

Calculate Hit

$$HIT\,(Link) := \text{How often is the given Link found in all projects}$$

Example:
In our 110 projects
$|beans.bean.class.X \Rightarrow \exists Entity^{FQN} = X| = 1000 \Rightarrow$

$$HIT\,(beans.bean.class) = 1000$$

Calculate Probability

$P(Link) :=$ Probability that the given Link **exists** in a project

Calculate Probability

$P\left(Link\right) :=$ Probability that the given Link **exists** in a project

Example:

$$P\left(beans.bean.class.X \Rightarrow \exists Entity^{FQN} = X\right) = \frac{HIT(beans.bean.class)}{SUP(beans.bean.class)}$$

Results

$\approx$ **110 Projects**

Java:

- $\approx$ 10 000 Entities
- $\approx$ 50 000 Methods
- $\approx$ 30 000 Annotations

XML:

- $\approx$ 5 000 Files
- $\approx$ 50 000 Tags
- $\approx$ 80 000 Attributes
- $\approx$ 70 000 Textual Contents

Results

$Probability > 90\%$; $Hit > 10$, $Framework = Spring$

Link	Type	Prob	Hit
beans.job.decision.decider	$Entity^{Name}$	97.44%	39
beans.job.step.tasklet.chunk.writer.bean.class	$Entity^{FQN}$	100%	25
beans.repositories.transaction-manager-ref	$Method^{Name}$	100%	23
beans.authentication-manager...data-source-ref	$Method^{Name}$	100%	13
beans.authorization-server.token-service-ref	$Method^{Name}$	91.67%	12
beans.flow.step.tasklet.ref	$Entity^{Name}$	100%	12
flow.var.class	$Entity^{FQN}$	91.67%	12
flow.var.name	$Entity^{Name}$	91.67%	12

Results

$Probability > 90\%$; $Hit > 10$, $Framework = Spring$

Link	Type	Prob	Hit
beans.job.decision.decider	$Entity^{Name}$	97.44%	39
beans.job.step.tasklet.chunk.writer.bean.class	$Entity^{FQN}$	100%	25
beans.repositories.transaction-manager-ref	$Method^{Name}$	100%	23
beans.authentication-manager...data-source-ref	$Method^{Name}$	100%	13
beans.authorization-server.token-service-ref	$Method^{Name}$	91.67%	12
beans.flow.step.tasklet.ref	$Entity^{Name}$	100%	12
flow.var.class	$Entity^{FQN}$	91.67%	12
flow.var.name	$Entity^{Name}$	91.67%	12

Where is $beans.bean.class \Rightarrow Entity.FQN$?

Frameworks

```
<bean id="oracleDataSource" class="org.springframework...">
    .....
</bean>

<bean id="oracleEntityManagerFactory" class="org.springframework...">
    <property name="dataSource" ref="oracleDataSource" />
    .....
</bean>

<bean id="transactionManager" class="org.springframework...">
    <property name="entityManagerFactory" ref="oracleEntityManagerFactory" />
</bean>
```

Calculate Hit

$HIT\,(Link)$:= How often is the given Link found in all projects **or** the definition refers to a framework

Calculate Probability

$P(Link)$:= Probability that the given Link exists in a project **or** the definition refers to a framework

Used Frameworks

SPRING	(.*)(springframework)(.*)
MAVEN	(.*)(maven\.apache\.org)(.*)
JAVAFX	(.*)(javafx\.com)(.*)
LIQUIBASE	(.*)(liquibase\.org)(.*)
WEBSPHERE	(.*)(websphere\.ibm\.com)(.*)
SOAPUI	(.*)(eviware\.com\/soapui)(.*)
GREPCONSOLE	(.*)(\/grepconsole)(.*)
JAVAEE_DEPLOYMENT	(.*)(java\.sun\.com\/xml\/ns\/javaee)(.*)
JAXB	(.*)(java\.sun\.com\/xml\/ns\/jaxb)(.*)
JAXWS	(.*)(java\.sun\.com\/xml\/ns\/jax-ws)(.*)
JAVA_SUN	(.*)(java\.sun\.com)(.*)
W3_ORG	(.*)(www\.w3\.org)(.*)
GOOGLE	(.*)(google\.com)(.*)
YAHOO	(.*)(yahoo\.com)(.*)
APACHE	(.*)(org\.apache)(.*)
MOZILLA	(.*)(www\.mozilla\.org)(.*)
NETBEANS	(.*)(www\.netbeans\.org)(.*)

Results

Probability $> 90\%$; *Hit* > 10, *Framework* $=$ *Spring*

Link	**Type**	**Prob**	**Hit**
beans.bean.class	$Entity^{FQN}$	90.56%	4 312
beans.repositories.base-package	$Package^{FQN}$	96.39%	83
beans.bean.property.list.bean	$Entity^{FQN}$	98.18%	55
beans.bean.property.bean.property.class	$Entity^{FQN}$	100%	51
beans.job.decision.decider	$Entity^{Name}$	97.44%	39
beans.job.step.tasklet.chunk.writer.bean.class	$Entity^{FQN}$	100%	25
beans.repositories.transaction-manager-ref	$Method^{Name}$	100%	23
beans.authentication-manager...data-source-ref	$Method^{Name}$	100%	13
beans.authorization-server.token-service-ref	$Method^{Name}$	91.67%	12
beans.flow.step.tasklet.ref	$Entity^{Name}$	100%	12
flow.var.class	$Entity^{FQN}$	91.67%	12
flow.var.name	$Entity^{Name}$	91.67%	12

Results

Probability $> 90\%$; *Hit* > 10, *Framework* $=$ *Spring*

Link	**Type**	**Prob**	**Hit**
beans.bean.class	$Entity^{FQN}$	90.56%	4 312
beans.repositories.base-package	$Package^{FQN}$	96.39%	83
beans.bean.property.list.bean	$Entity^{FQN}$	98.18%	55
beans.bean.property.bean.property.class	$Entity^{FQN}$	100%	51
beans.job.decision.decider	$Entity^{Name}$	97.44%	39
beans.job.step.tasklet.chunk.writer.bean.class	$Entity^{FQN}$	100%	25
beans.repositories.transaction-manager-ref	$Method^{Name}$	100%	23
beans.authentication-manager...data-source-ref	$Method^{Name}$	100%	13
beans.authorization-server.token-service-ref	$Method^{Name}$	91.67%	12
beans.flow.step.tasklet.ref	$Entity^{Name}$	100%	12
flow.var.class	$Entity^{FQN}$	91.67%	12
flow.var.name	$Entity^{Name}$	91.67%	12

beans.bean.class $\Rightarrow$ *Entity.FQN* is now under the result

Next Steps

Find complex rules

Next Steps

Find complex rules

- Association rule learning *(FP-growth algorithm)*

Next Steps

Find complex rules

- Association rule learning *(FP-growth algorithm)*
- Pattern mining in a model-graph

Overview

Overview

Software is not only a source code

Programming Language

XML Configuration Files

Frameworks

Overview

Software is not only a source code

Programming Language

XML Configuration Files

Frameworks

Motivation

We will

- analyse **both**, source code and XML files,
- **link** them together in one model,
- discover **patterns**

Overview

Software is not only a source code

Programming Language

XML Configuration Files

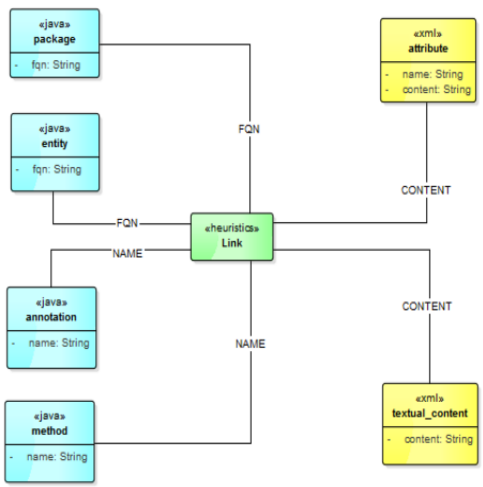
Frameworks

Motivation

We will

- analyse **both**, source code and XML files,
- **link** them together in one model,
- discover **patterns**

Link Java and XML Model



Overview

Software is not only a source code

Programming Language

XML Configuration Files

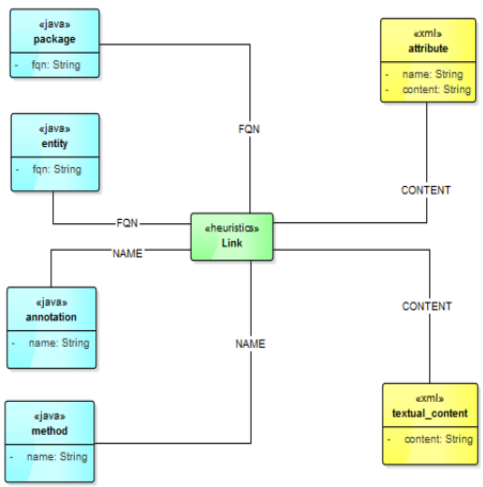
Frameworks

Motivation

We will

- analyse **both**, source code and XML files,
- **link** them together in one model,
- discover **patterns**

Link Java and XML Model



Probability $> 90\%$; *Hit* > 10, *Framework* $=$ *Spring*

Link	Type	Prob	Hit
beans.bean.class	$Entity^{FQN}$	90.56%	4 312
beans.repositories.base-package	$Package^{FQN}$	96.39%	83
beans.bean.property.list.bean	$Entity^{FQN}$	98.18%	55
beans.bean.property.bean.property.class	$Entity^{FQN}$	100%	51
beans.job.decision.decider	$Entity^{Name}$	97.44%	39
beans.job.step.tasklet.chunk.writer.bean.class	$Entity^{FQN}$	100%	25
beans.repositories.transaction-manager-ref	$Method^{Name}$	100%	23
beans.authentication-manager...data-source-ref	$Method^{Name}$	100%	13
beans.authorization-server.token-service-ref	$Method^{Name}$	91.67%	12
beans.flow.step.tasklet.ref	$Entity^{Name}$	100%	12
flow.var.class	$Entity^{FQN}$	91.67%	12
flow.var.name	$Entity^{Name}$	91.67%	12