

# SrcML tool 生成 AST

## 1. 安装 SrcML

在“<https://www.srcml.org/#download>”下载 SrcML 的安装程序(谷歌浏览器可能跳不出来下载页面，用火狐可以下载)。

### srcML v1.0.0

The latest release (January 2020) - installers and executables. Downloads include the command line interface client, *srcml*, and the srcML library, *libsrcml*. Linux downloads have a separate installer to develop applications that use the libsrcml library. Let us know if you need an installer for a platform not listed here.

Windows	
Windows 10	<a href="#">exe</a>
macOS	
Catalina 10.15	<a href="#">pkg</a>
Mojave 10.14	<a href="#">pkg</a>
Linux	<div>Client</div> <div>Develop</div>
Ubuntu 20.04	<a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a> <a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a>
Ubuntu 19.10	<a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a> <a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a>
Ubuntu 19.04	<a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a> <a href="#">deb</a> <a href="#">tar.gz</a> <a href="#">tar.bz2</a>

下载完之后点击运行，用默认设置安装就行了。

## 2. 使用 SrcML

参考“<https://www.srcml.org/tutorials/creating-srcml.html>”的介绍，在命令行输入对应的指令即可。

### Creating srcML

Let's start off by making a simple srcML unit. We have a C++ file called `rotate.cpp`. To transform this file to the srcML format, execute `srcml` on the command line with the path to the file.

```
$ srcml rotate.cpp
```

Outputting to a file can be accomplished with the `-o` option, or by piping standard output to a file. For example, the following command will output the resulting srcML format to a `rotate.xml` file. The `xml` extension is important in order to parse it as a srcML file later, and all srcML is valid XML.

```
$ srcml rotate.cpp -o rotate.xml
```

### 3. output error

出现 output error 是因为路径中有中文，但是不影响 xml 文件生成。

```
output error : string is not in UTF-8
output error : string is not in UTF-8
```

如果直接用浏览器打开生成的 xml 文件，会出现错误。

#### This page contains the following errors:

error on line 2 at column 1: Encoding error

Below is a rendering of the page up to the first error.

需要先用 eclipse 或者其他 IDE 打开 xml 文件，把乱码部分改成对应的中文，然后就可以用浏览器正常打开了。

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<unit xmlns="http://www.srcML.org/srcML/src" revision="1.0.0" language="Java" filename="C:\Users\用户\Desktop\code2seq-master\Java\src\main\java\com\srcML\code2seq\Code2SeqTest.java">
  <comment type="line">// void fooBar() {</comment>
  <comment type="line">// System.out.println("http://github.com");</comment>
  <comment type="line">// }</comment>
  <comment type="line">// void thisIsName() {</comment>
  <comment type="line">// System.out.println("a");</comment>
  <comment type="line">// }</comment>
  <comment type="line">//</comment>
  <class>class <name>Test</name> <block>{
    <function><type><specifier>public</specifier> <name>String</name></type> <name>extractFor</name><parameter_list>(<parameter><decl><
      <expr_stmt><expr><call><name>LOG</name><operator>.</operator><name>debug</name></name><argument_list>(<argument><expr><li>
        <return>return <expr><call><name>requests</name><operator>.</operator><name>remove</name></name><argument_list>(<argument
      </block_content></block></function>
  }</block></class></unit>
```

This XML file does not appear to have any style information associated with it. The tree is shown below.

```
▼<unit xmlns="http://www.srcML.org/srcML/src" revision="1.0.0" language="Java" filename="C:\Users\
\Desktop\code2seq-master\JavaExtractor\JPredict\src\main\java\Test.java">
  <comment type="line">//class Test {</comment>
  <comment type="line">// void fooBar() {</comment>
  <comment type="line">// System.out.println("http://github.com");</comment>
  <comment type="line">// }</comment>
  <comment type="line">// void thisIsName() {</comment>
  <comment type="line">// System.out.println("a");</comment>
  <comment type="line">// }</comment>
  <comment type="line">//}</comment>
  ▼<class>
    class
    <name>Test</name>
    ▼<block>
      {
      ▼<function>
        ▼<type>
          <specifier>public</specifier>
          <name>String</name>
        </type>
        <name>extractFor</name>
        ▼<parameter_list>
          (
          ▼<parameter>
            ▼<decl>
              ▼<type>
                <name>Integer</name>
              </type>
              <name>id</name>
            </decl>
          </parameter>
        )
      )
    )
  )

```

#### 4. 读取 xml 文件

在 maven 项目中添加 dom4j 的依赖，并更新项目导入 ja 包。

```
<dependency>
```

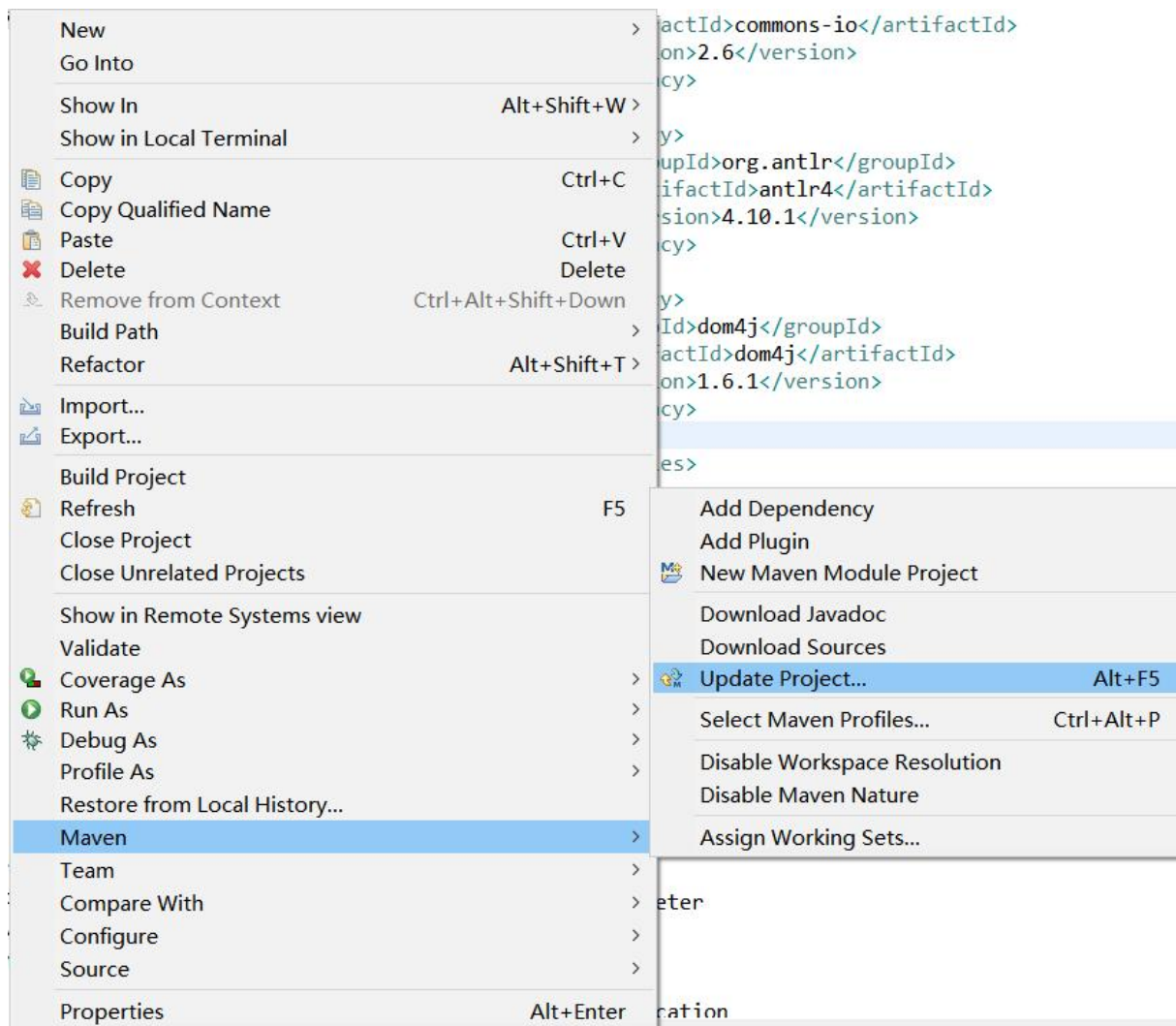
```
  <groupId>dom4j</groupId>
```

```
  <artifactId>dom4j</artifactId>
```

```
  <version>1.6.1</version>
```

```
</dependency>
```

右键项目—>【Maven】—>【Update Project】，下载 jar 包。



代码:

```
public class XmlToAST {

    public static void main(String[] args) {

        try {

            SAXReader reader = new SAXReader();

            Document document = reader.read(new
File("C:\\Users\\xxx\\Desktop\\AST\\Test.xml"));

            Element rootElement = document.getRootElement();

            Element element =rootElement.element("class");
```

```

        printAST(element);

    } catch (DocumentException e) {
        e.printStackTrace();
    }


}

public static void printAST(Element root)
{
    System.out.print("(" + root.getName());
    List<Element> elements = root.elements();
    if(elements.size() == 0) System.out.print("(" + root.getText() + ")");
    else
    {
        for(Element element: elements){
            printAST(element);
        }
    }
    System.out.print(")");
}
}

```

## 5. 节点的 label

查看 <https://www.srcml.org/doc/srcMLElements.html>



<srcML>

HOME

DOWNLOAD

ABOUT

TOOLS

DOCUMENTATION

TUTORIALS

SUPPORT


COMMUNITY

srcML 1.0.0 Elements

✓ link to the element documentation for that language

Element	C	C++	C#	Java
<alignas>	✓	✓		
<alignof>	✓	✓		
<annotation>				✓
<annotation_defn>				✓
<asm>	✓	✓		
<assert type="static">	✓	✓		
<assert>				✓
<association>	✓			
<association_list>	✓			
<atomic>	✓			
<attribute>		✓	✓	
<block>	✓	✓	✓	✓
<block_content>	✓	✓	✓	✓
<break>	✓	✓	✓	✓

点击 Documentation—>对应的语言是对某种语言的节点的介绍



<srcML>

HOME

DOWNLOAD

ABOUT

TOOLS

DOCUMENTATION

TUTORIALS

SUPPORT


COMMUNITY

srcML 1.0.0 Elements

C++C#JavaCPPELEMENTSSUPPORT LEVELfor that language

Element	C	C++	C#	Java
<alignas>	✓	✓		
<alignof>	✓	✓		
<annotation>				✓
<annotation_defn>				✓
<asm>	✓	✓		
<assert type="static">	✓	✓		
<assert>				✓
<association>	✓			
<association_list>	✓			

以 java 为例



<srcML>

HOME

DOWNLOAD

ABOUT

TOOLS

DOCUMENTATION

TUTORIALS

SUPPORT

COMMUNITY

srcML v1.0.0 Java Documentation

Base Elements	XML Element	Subelements
comment	<comment type="block line" format="javadoc doxygen">	
escape	<escape char="0xXX">	
literal	<literal type="number char string boolean null">	
name	<name>	<name>, <operator>

Statements	XML Element	Subelements
assert statement	<assert>	<expr>
block	<block>	<block_content>
break statement	<break>	<name>
case statement	<case>	<expr>
continue statement	<continue>	<name>
default statement	<default>	
do while statement	<do>	<block>, <condition>