

# Automated detection and resolution of legal cross references

Approach and a study of Luxembourg's legislation

Morayo Adedjouma<sup>1,2</sup> Mehrdad Sabetzadeh<sup>1</sup> Lionel Briand<sup>1</sup>

<sup>1</sup> University of Luxembourg, Luxembourg

<sup>2</sup> McMaster University, Canada

# Luxembourg

- Population of 530,000
- The only Grand Duchy in existence!
- Three official languages



- Language of administration and law is French

# How did this work come about?

- Collaboration with Government of Luxembourg



LE GOUVERNEMENT  
DU GRAND-DUCHÉ DE LUXEMBOURG

- **CTIE**: Government's IT Centre
- **ACD**: Tax Administration Department
- New tax system under development
- System needs to be **compliant** with the law

# Relating system requirements to legal texts

**Art. 2** Individuals are considered resident taxpayers if they have either their fiscal or habitual residence in the Grand Duchy. Individuals are considered non-resident taxpayers if they neither have their fiscal nor their habitual residence in the Grand Duchy, but have a local income within the meaning of Art.156.

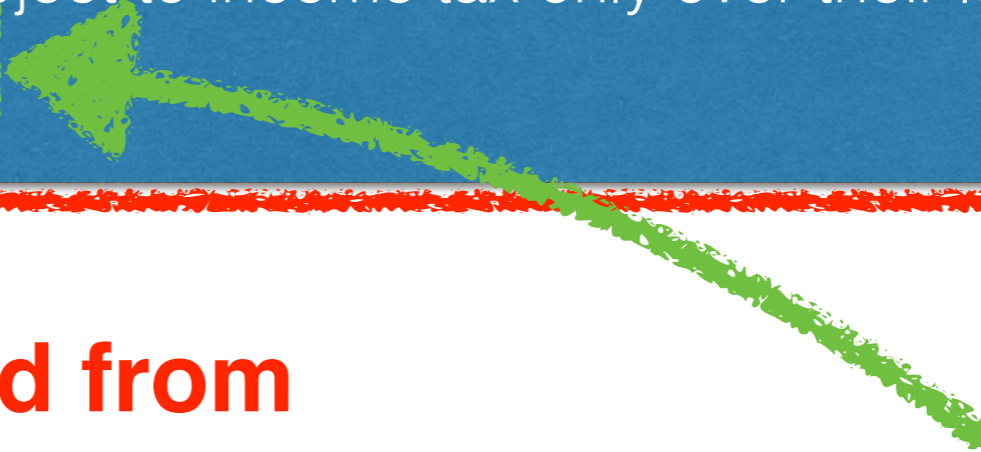
Resident taxpayers are subject to income tax over both local and foreign income.

Non-resident taxpayers are subject to income tax only over their local income within the meaning of Art.156.



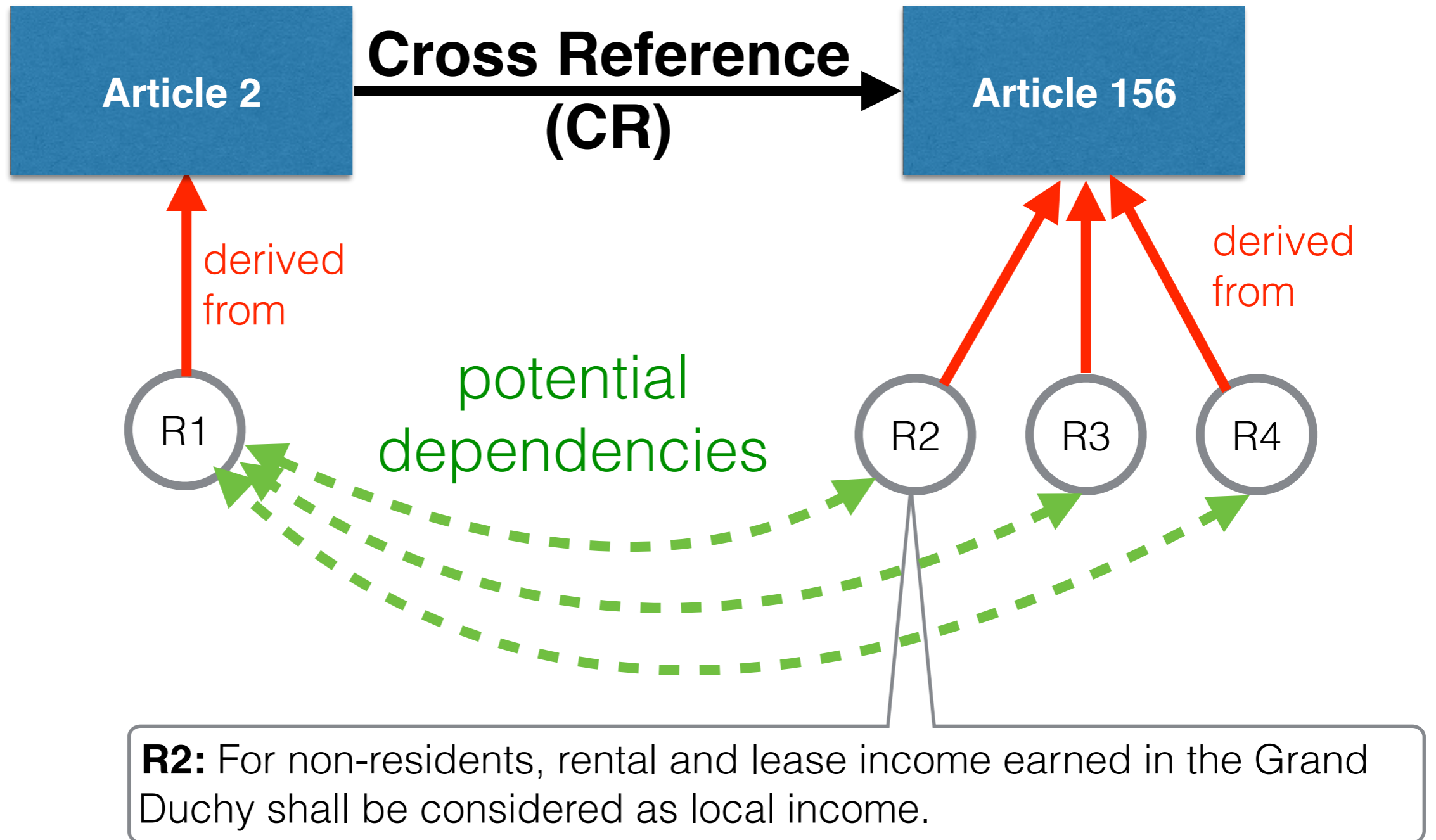
**derived from**

**R1:** The system shall levy taxes on non-residents' local income as per the annual tax scale.



**Cross  
Reference  
to another  
article**

# Requirement dependencies



# Why should RE care about CRs?

- CRs important for requirements change analysis  
**[Ghanavati et al., 2014]**
- CRs entail (among other things) exceptions, constraints, and priorities **[Maxwell et al., 2012]**
- Ignoring or misunderstanding CRs leads to non-compliance in software **[Maxwell et al., 2013]**

# CR detection and resolution

- Automatically identify and link CRs to targets

## Art. 8.

1 Lorsqu'un contribuable a son domicile fiscal au Grand  
une ou plusieurs habitations secondaires, son revenu  
somme égale au minimum à cinq et au maximum à dix  
de la ou des habitations secondaires, à moins que le c  
raison de son revenu au sens de l' [article 7](#) .

2 Le ministre des finances arrêtera un multiplicateur  
eu égard à la relation normalement existante entre le r

3 Lorsque le revenu forfaitaire visé à l' [alinéa 1er](#) est re  
dispositions des [articles 126 à 134ter](#) ne sont pas appl

[Article 126](#)  
[Article 127](#)  
[Article 127bis](#)  
[Article 127ter](#)  
[Article 127quater](#)  
[Article 128](#)  
[Article 128bis](#)  
[Article 129](#)  
[Article 129a](#)  
[Article 129b](#)  
[Article 129c](#)  
[Article 130](#)  
[Article 131](#)  
[Article 132](#)  
[Article 133](#)  
[Article 134](#)  
[Article 134bis](#)  
[Article 134ter](#)

[Alinea 1](#)

[alinéa qui précède](#)

# Related work

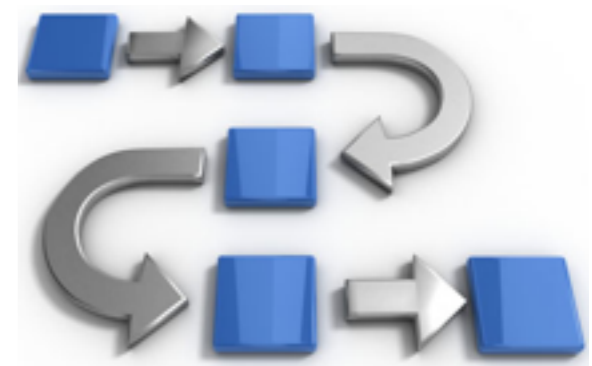
- CR detection and resolution:  
a well-trodden path
- Dutch legislation  
**[de Maat et al., 2006]**
- HIPAA Privacy Rule  
**[Kiyavitskaya et al., 2008]**
- US regulations  
**[Breux, 2009]**





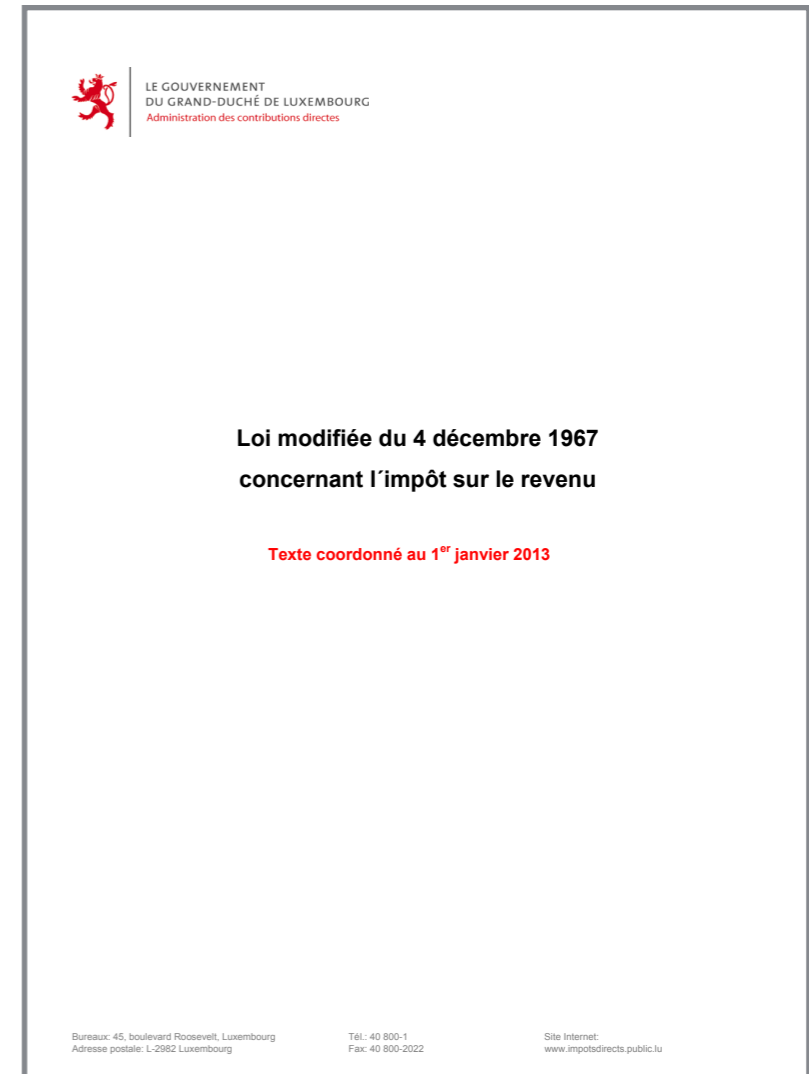
# What's new?

- More thorough treatment of CR patterns
- Systematic resolution process
- Improved automation



# CR patterns

- Grounded Theory study of Luxembourg's Income Tax Law
  - first drafted in 1967;  
189 pages; 236 articles;  
767 paragraphs
- All 1223 CRs analysed

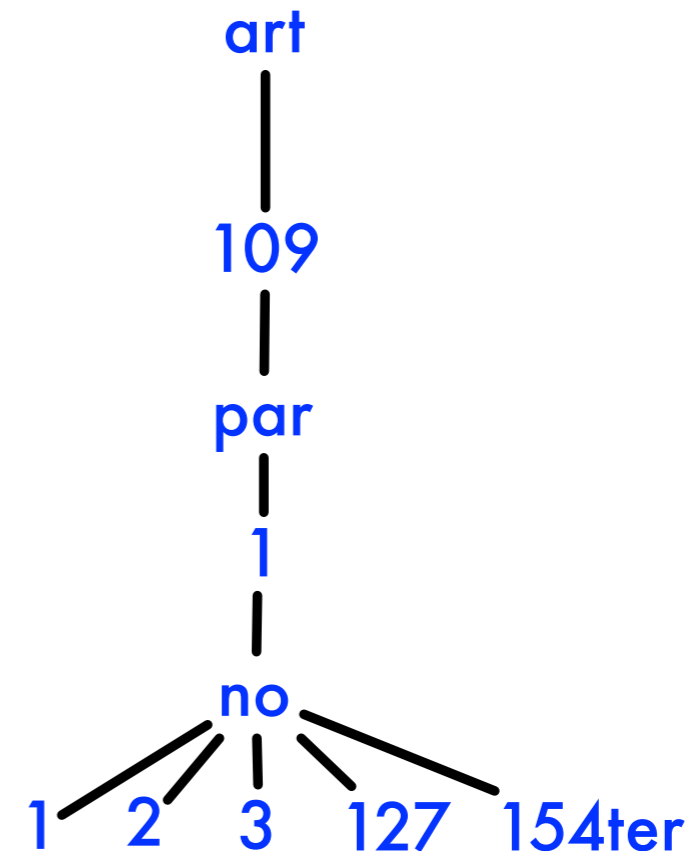
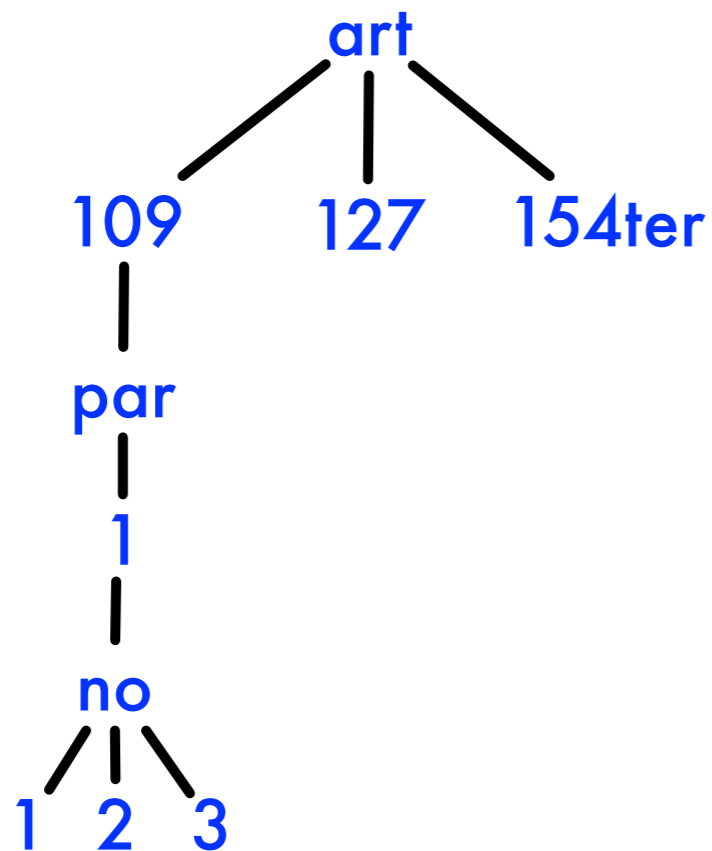


# What did we learn?

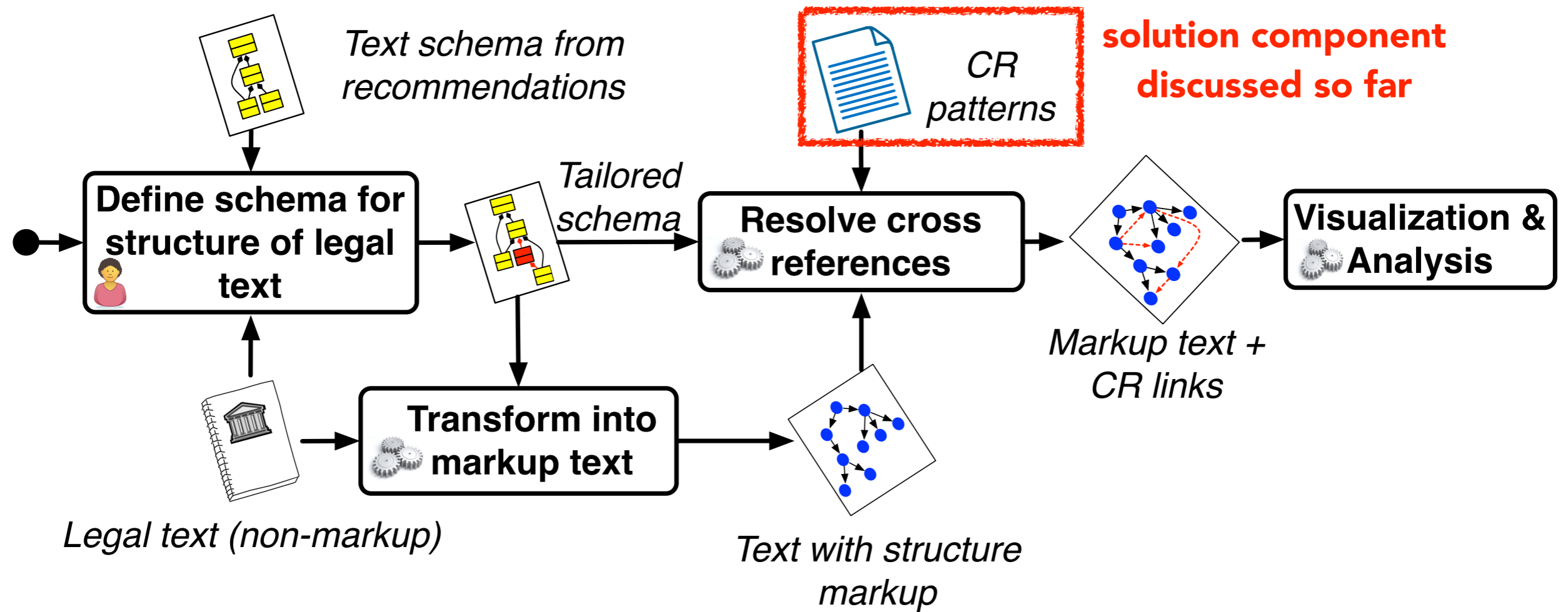
- CR patterns carry over across languages
  - thus, opportunities for reuse
- All previously-known patterns were observed
- New patterns were identified
- ... among which some are ambiguous

# Example of ambiguity

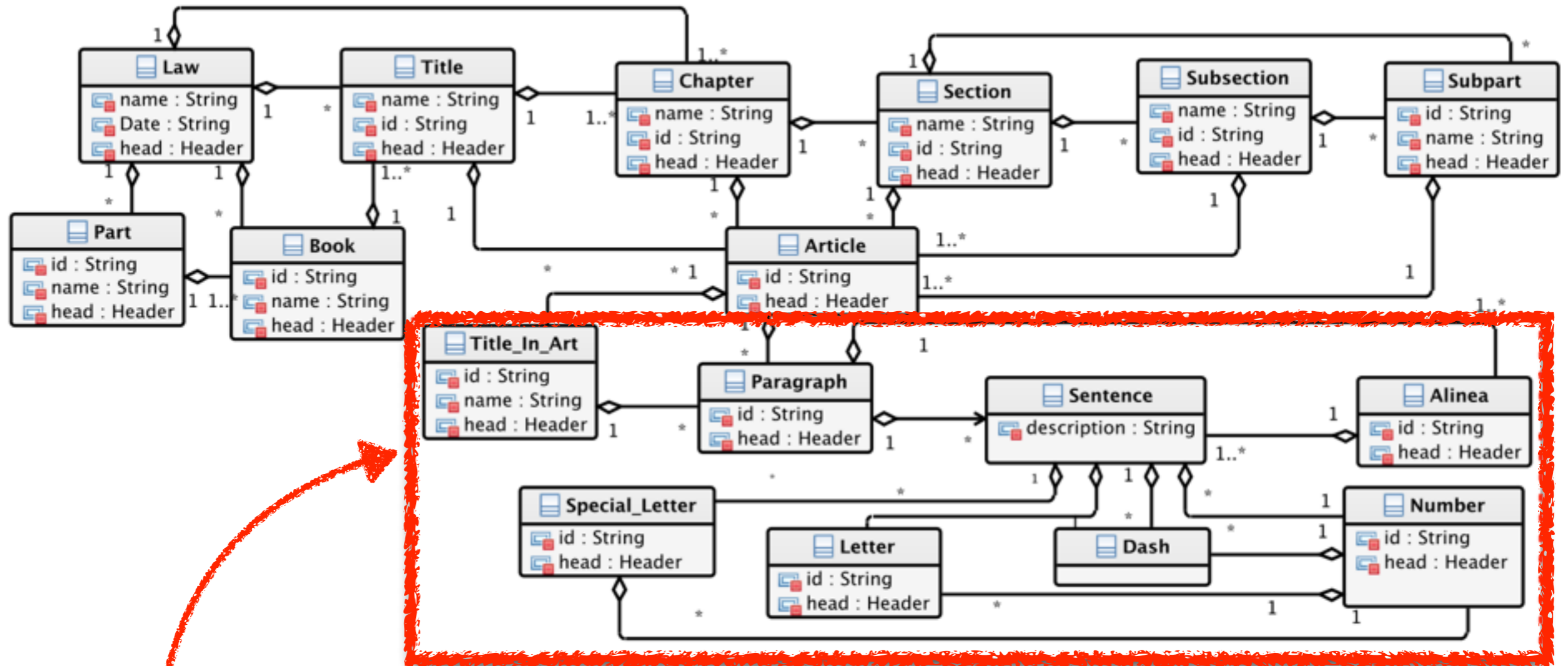
**articles 109, 1st paragraph, numbers 1 to 3, 127 and 154ter**



# Approach



# Text schema for the tax law



New or deviating from guidelines

# Markup scripts derived from the text schema

Whatever is above or unrelated to Articles

## Rule: MarkArticleSegment

```
(({ArticleHead}):start
```

```
{LawHead} | {PartHead} | {BookHead} | {TitleHead} |  
{ChapterHead} | {SectionHead} | {SubsectionHead} |  
{SubpartHead} | {ArticleHead} | {EOD}):end)
```

```
:ref -->:ref.ArticleSegment= {}
```

## Rule: MarkArticleHead

```
(({Split})+ ({Token.string=="Art"} | {Token.string=="Article"}))
```

```
{Token.string=="."}? {Token.kind=="alphanumeric"}
```

```
{Token.string=="."}
```

```
:ref -->:ref.ArticleHead= {}
```

# Example markup

Article head

Art. 8. (1) Les propositions élaborées par le centre concernant la solution intégrée des problèmes d'informatique communs à l'ensemble ou à certaines administrations pourront, après consultation obligatoire du comité visé à l'article 7, être déclarées par le Gouvernement en conseil d'application obligatoire pour tous les services intéressés.  
(2) Les contestations pouvant s'élever en matière informatique entre deux ou plusieurs administrations de l'Etat ou entre une administration et le centre sont tranchées par le Gouvernement en conseil sur avis préalable du comité visé à l'article 7.

Article segment

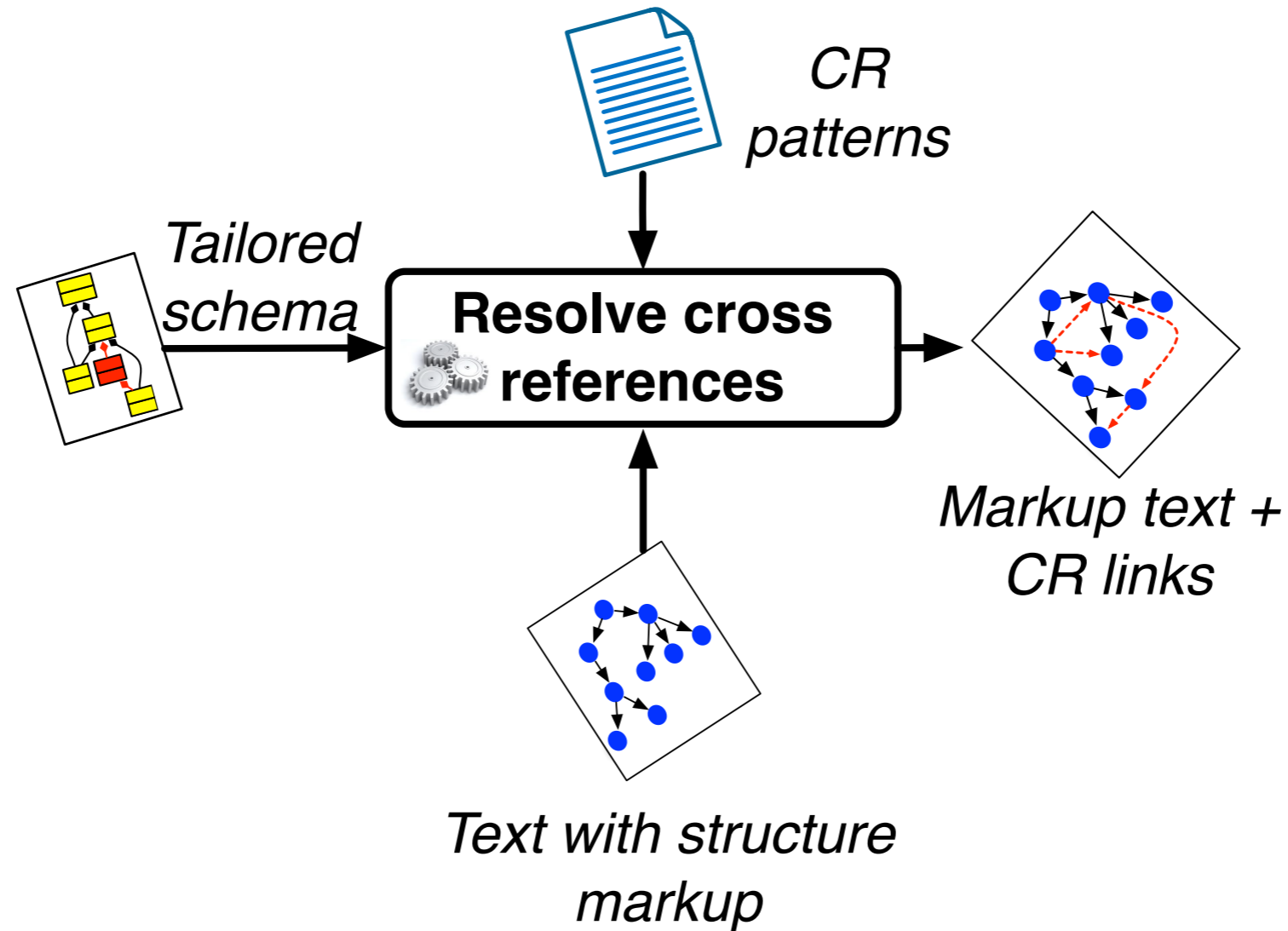
- Markup generated all the way down to sentences

Paragraphs

Art. 8. (1) Les propositions élaborées par le centre concernant la solution intégrée des problèmes d'informatique communs à l'ensemble ou à certaines administrations pourront, après consultation obligatoire du comité visé à l'article 7, être déclarées par le Gouvernement en conseil d'application obligatoire pour tous les services intéressés.  
(2) Les contestations pouvant s'élever en matière informatique entre deux ou plusieurs administrations de l'Etat ou entre une administration et le centre sont tranchées par le Gouvernement en conseil sur avis préalable du comité visé à l'article 7.



# CR Resolution



- Resolution algorithm discussed in the paper

# CR visualization and analysis

- Legal text navigation
- Trace detection

**Art. 8.**

1 Lorsqu'un contribuable a son domicile fiscal au Grand-Duché de Luxembourg et qu'il possède une ou plusieurs habitations secondaires, son revenu est imposable de manière forfaitaire à une somme égale au minimum à cinq et au maximum à dix pour cent du revenu brut ou la valeur locative brute de la ou des habitations secondaires, à moins que le contribuable ne demande à être imposé à raison de son revenu au sens de l' [article 7](#) .

2 Le ministre des finances arrêtera un multiplicateur à appliquer au revenu forfaitaire en fonction de la relation normalement existante entre le revenu forfaitaire et le loyer brut d'habitation.

3 Lorsque le revenu forfaitaire visé à l' [alinéa 1er](#) est révisé, les dispositions des [articles 126 à 134ter](#) ne sont pas applicables.



The screenshot shows a legal article with a navigation menu on the right. The menu lists articles from 126 to 134ter. A red arrow points from the 'articles 126 à 134ter' link in the text to the corresponding menu item. A red box highlights the 'articles 126 à 134ter' link in the text.

Title I::Chapter IV::Section I::Article 8	<b>4 citations:</b> <ul style="list-style-type: none"><li>• Title I::Chapter III::Article 3bis::Paragraph 2</li><li>• Title I::Chapter IV::Section I::Article 8::Paragraph 2</li><li>• Title I::Chapter IV::Section I::Article 8::Paragraph 3</li><li>• Title I::Chapter VIII::Section II::Article 141::Paragraph 1</li></ul>
---	---

- Circularity analysis

# Completeness of CR patterns

- Studied 164 randomly-selected pages from a corpus of 13 legal texts (1640 pages)
  - Oldest text drafted in 1808; newest in 2011
- 1852 CRs in the selected pages
  - Correctly identified 1813 CRs
  - Detected 38 CRs only partially
  - Missed 1 CR completely
  - 5 false positives

Detection accuracy

**Precision:** 99.7%

**Recall:** 97.9%

# Accuracy of resolution

- Resolution attempted for tax law's 928 **internal** CRs
- 45 CRs (5%) marked for manual resolution
- 874 CRs (94%) automatically resolved (1736 links)
- 9 failure-to-resolve warnings
  - 8 due to anomalies in the text
  - 1 false positive due to limitations of the approach

Resolution accuracy  
**over CR links**

**Precision:** 99.9%

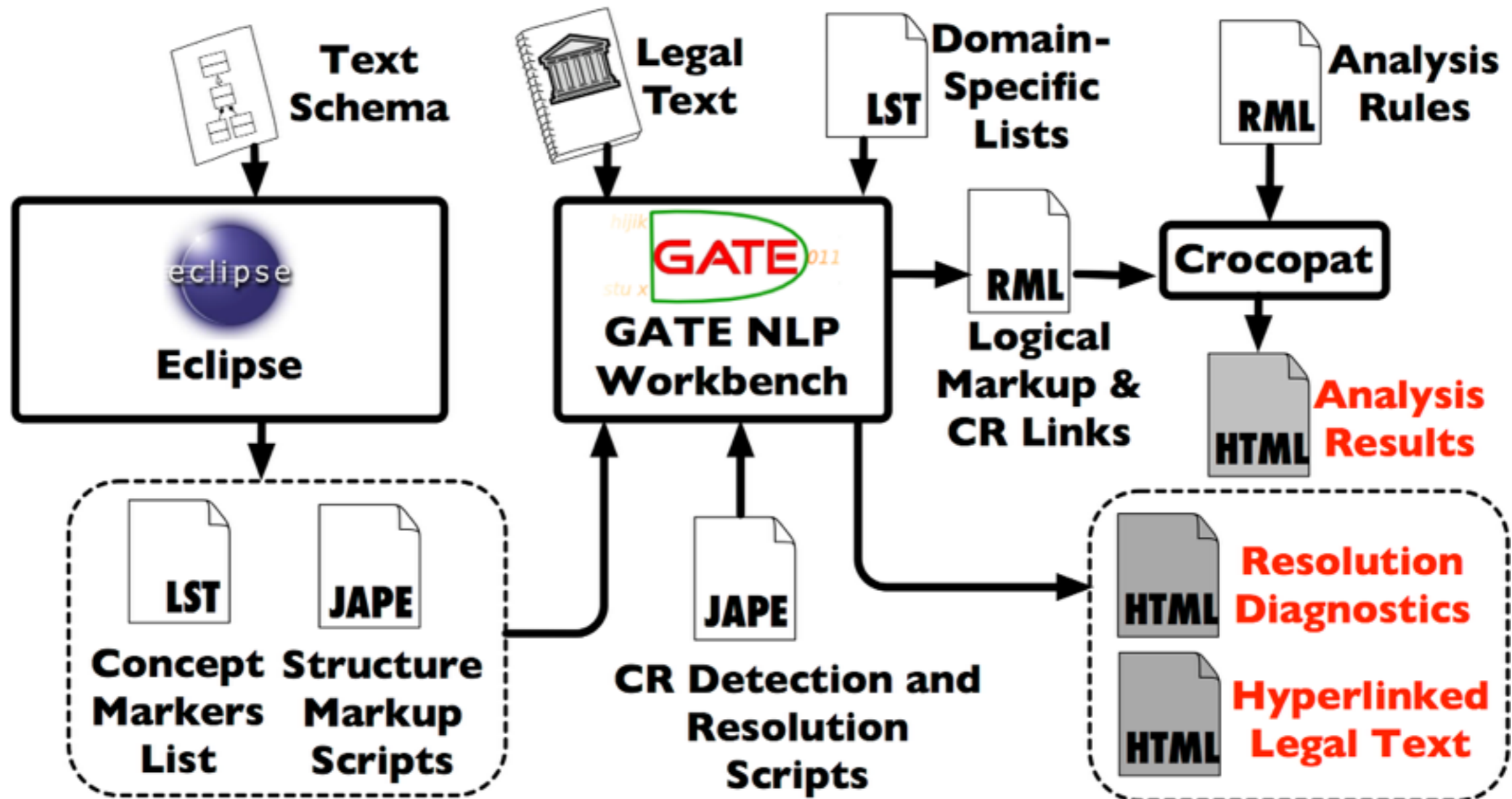
**Recall:** 97.5%

# Scalability

- Results for the income tax law (189 pages):

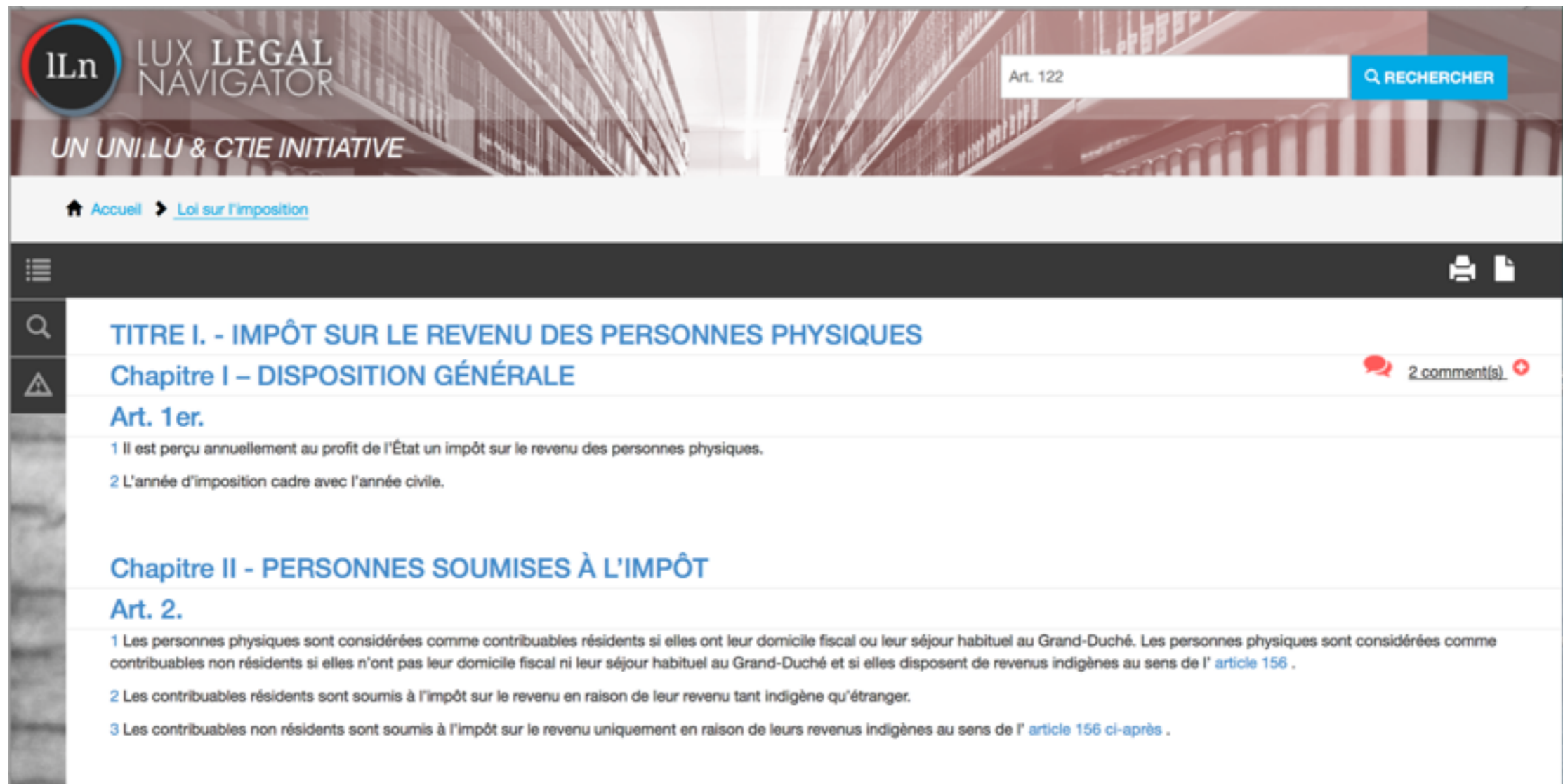
Structural markup generation + CR identification	34 sec
Resolution, generation of hyperlinked legal text, traceability matrix and diagnostics	290 sec
<b>Total</b>	<b>Approx. 5 minutes</b>

# Tool Support



# A beneficial byproduct

- Automated legal portal generation



The screenshot displays the LUX LEGAL NAVIGATOR website interface. At the top left, the logo 'ILn LUX LEGAL NAVIGATOR' is visible, along with the text 'UN UNI.LU & CTIE INITIATIVE'. A search bar on the right contains 'Art. 122' and a 'RECHERCHER' button. Below the header, a breadcrumb trail shows 'Accueil > Loi sur l'imposition'. The main content area is titled 'TITRE I. - IMPÔT SUR LE REVENU DES PERSONNES PHYSIQUES' and 'Chapitre I - DISPOSITION GÉNÉRALE'. It lists 'Art. 1er.' with two paragraphs: '1 Il est perçu annuellement au profit de l'État un impôt sur le revenu des personnes physiques.' and '2 L'année d'imposition cadre avec l'année civile.' Below this, 'Chapitre II - PERSONNES SOUMISES À L'IMPÔT' is shown, followed by 'Art. 2.' with three paragraphs detailing tax residency rules. A sidebar on the left contains navigation icons, and a top right corner has print and share icons. A comment count '2 comment(s)' is visible next to the chapter title.

*webpage courtesy of CTIE*

# Limitations

- All CR types detected but only internal ones are currently resolved
- Limited to legislative documents
  - Regulations, circular letters, directives, and parliamentary proceedings not covered



# What's next?

- Generation of semantic annotations for CRs
- Change impact analysis for legal requirements

# Take away

- Example of a transition from research to a palatable solution to the industry
- Lesson learned from the transition:  
There is many a slip between the  
cup and the lip!



# Acknowledgements

- **CTIE:** Ludwig Balmer, Marc Blau, Yves Boland, Michaël Masseroni, Gautier Barrère
- **ACD:** Thierry Prommenschenkel, Guy Heintz, Vincent D'Angelo
- **SnT:** Domenico Bianculli