

Intelligent Regulation Splitter

Background description

Regulatory documents follow a similar yet unique structure in each case. On a high level they are multi-level text structures. They (can) consist of Sections, Parts that contain Articles, which can be split into Sub-articles etc.

Each level has some sort of identifier.

Typical document structure example:

Case 1: with Section and Subsection

Regulation

```
|-- Section: <title>
|   |-- Subsection: <title>
|       |-- Article <number>: <title>
|           |-- Item [ref=1]
|           |-- Item [ref=1(a)]
|       |-- Article <number>: <title>
```

Case 2: no Section/Subsection

Regulation

```
|-- Article <number>: <title>
|   |-- Item [ref=1]
|   |-- Item [ref=""]
|-- Article <number>: <title> (no items, the Article is indivisible).
```

Task

Build a tool that ingests a regulation document (HTML, PDF, plain text) and breaks it down into a document graph with:

- * sections
- * articles
- * paragraphs
- * **references between them** etc.

Requirements

- **LLMs/other AI tech is allowed and encouraged**

- Works across jurisdictions and numbering styles (e.g., Article 2, Art. 2º, Section 1, 1(a)).
- Actual regulation text is preserved verbatim (no rephrasing).
- Robust to noisy HTML/PDF (ignores headers/footers/footnotes/page numbers).
- Bonus if can reason through crawling, for example if given a link https://www.kslegislature.gov/li_2012/b2011_12/statute/009_000_0000_chapter/009_005_0000_article/, should understand it needs to collect the structure from underlying links.
- Bonus: language-agnostic.
- Bonus: supports scanned PDFs via OCR.
- Bonus: human in the loop support.

Acceptance

- Provide a way of running the solution against a URL.
- We will run your solution on the examples and other documents and check article/item boundaries, titles/refs, and ordering, for example:
 - <https://www.mastercard.com/content/dam/mccom/shared/business/support/rules-pdfs/mastercard-rules.pdf>
 - <https://www.scb.gov.bs/wp-content/uploads/2024/07/Gazetted-Digital-Assets-and-Registered-Exchanges-Act-2024.pdf>
 - <https://sdlegislature.gov/Statutes/51A-17>
 - <https://rules.sos.ga.gov/gac/80-3>

Deliverables

Source code, tests, and a README with setup/run instructions.

Examples

EU example (MiCA)

JSON

```
{
  "regulation": {
    "title": "Regulation (EU) 2023/1114 (MiCA)",
    "url":
      "https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32023R1114"
  },
  "articles": [
    {
      "title": "Article 2 - Scope",
      "content": "Article 2\nScope\n1. This Regulation applies to natural and legal persons and certain other undertakings that are engaged in the issuance, offer to the public and admission to trading of crypto-assets or that provide services related to crypto-assets in the Union.\n\n2. This Regulation does not apply to:\n(a) persons who provide crypto-asset services exclusively for their parent companies, for their own subsidiaries or for other subsidiaries of their parent companies;\n(b) a liquidator or an administrator acting in the course of an insolvency procedure, except for the purposes of Article 47;\n(c) the ECB, central banks of the Member States when acting in their capacity as monetary authorities, or other public authorities of the Member States;\n(d) the European Investment Bank and its subsidiaries;\n(e) the European Financial Stability Facility and the European Stability Mechanism;\n(f) public international organisations.\n\n3. This Regulation does not apply to crypto-assets that are unique and not fungible with other crypto-assets.\n\n4. This Regulation does not apply to crypto-assets that qualify as one or more of the following:\n(a) financial instruments;\n(b) deposits, including structured deposits;\n(c) funds, except if they qualify as e-money tokens;\n(d) securitisation positions in the context of a securitisation as defined in Article 2, point (1), of Regulation (EU) 2017/2402;\n(e) non-life or life insurance products falling within the classes of insurance listed in Annexes I and II to Directive 2009/138/EC of the European Parliament and of the Council (27) or reinsurance and retrocession contracts referred to in that Directive;\n(f) pension products that, under national law, are recognised as having the primary purpose of providing the investor with an income in retirement and that entitle the investor to certain benefits;\n(g) officially recognised occupational pension schemes falling within the scope of Directive (EU) 2016/2341 of the European Parliament and of the Council (28) or Directive 2009/138/EC;\n(h) individual pension products for which a financial contribution from the employer is required by national law and where the
```

employer or the employee has no choice as to the pension product or provider;\n(i) a pan-European Personal Pension Product as defined in Article 2, point (2), of Regulation (EU) 2019/1238 of the European Parliament and of the Council (29);\n(j) social security schemes covered by Regulations (EC) No 883/2004 (30) and (EC) No 987/2009 of the European Parliament and of the Council (31).\n\n5. By 30 December 2024, ESMA shall, for the purposes of paragraph 4, point (a), of this Article issue guidelines in accordance with Article 16 of Regulation (EU) No 1095/2010 on the conditions and criteria for the qualification of crypto-assets as financial instruments.\n\n6. This Regulation shall be without prejudice to Regulation (EU) No 1024/2013.",

"items": [

{ "ref": "1", "content": "1. This Regulation applies to natural and legal persons and certain other undertakings that are engaged in the issuance, offer to the public and admission to trading of crypto-assets or that provide services related to crypto-assets in the Union." },

{ "ref": "2", "content": "2. This Regulation does not apply to:\n(a) persons who provide crypto-asset services exclusively for their parent companies, for their own subsidiaries or for other subsidiaries of their parent companies;\n(b) a liquidator or an administrator acting in the course of an insolvency procedure, except for the purposes of Article 47;\n(c) the ECB, central banks of the Member States when acting in their capacity as monetary authorities, or other public authorities of the Member States;\n(d) the European Investment Bank and its subsidiaries;\n(e) the European Financial Stability Facility and the European Stability Mechanism;\n(f) public international organisations." },

{ "ref": "3", "content": "3. This Regulation does not apply to crypto-assets that are unique and not fungible with other crypto-assets." },

{ "ref": "4", "content": "4. This Regulation does not apply to crypto-assets that qualify as one or more of the following:\n(a) financial instruments;\n(b) deposits, including structured deposits;\n(c) funds, except if they qualify as e-money tokens;\n(d) securitisation positions in the context of a securitisation as defined in Article 2, point (1), of Regulation (EU) 2017/2402;\n(e) non-life or life insurance products falling within the classes of insurance listed in Annexes I and II to Directive 2009/138/EC of the European Parliament and of the Council (27) or reinsurance and retrocession contracts referred to in that Directive;\n(f) pension products that, under national law, are recognised as having the primary purpose of providing the investor with an income in retirement and that entitle the investor to certain benefits;\n(g) officially recognised occupational pension schemes falling within the scope of Directive (EU) 2016/2341 of the European Parliament and of the Council (28) or Directive 2009/138/EC;\n(h) individual pension products for which a financial contribution from the employer is required by national law and where the employer or the employee has no choice as to the pension product or provider;\n(i) a pan-European Personal Pension Product as defined in Article

2, point (2), of Regulation (EU) 2019/1238 of the European Parliament and of the Council (29);\n(j) social security schemes covered by Regulations (EC) No 883/2004 (30) and (EC) No 987/2009 of the European Parliament and of the Council (31)."},

{"ref": "5", "content": "5. By 30 December 2024, ESMA shall, for the purposes of paragraph 4, point (a), of this Article issue guidelines in accordance with Article 16 of Regulation (EU) No 1095/2010 on the conditions and criteria for the qualification of crypto-assets as financial instruments."},

{"ref": "6", "content": "6. This Regulation shall be without prejudice to Regulation (EU) No 1024/2013."}

],

"path": []

},

{

"title": "Article 4 - Offers to the public of crypto-assets other than asset-referenced tokens or e-money tokens",

"content": "Article 4\nOffers to the public of crypto-assets other than asset-referenced tokens or e-money tokens\n1. A person shall not make an offer to the public of a crypto-asset other than an asset-referenced token or e-money token in the Union unless that person: ...",

"items": [

{"ref": "1", "content": "1. A person shall not make an offer to the public of a crypto-asset other than an asset-referenced token or e-money token in the Union unless that person:\n(a) is a legal person;\n(b) has drawn up a crypto-asset white paper in respect of that crypto-asset in accordance with Article 6;\n(c) has notified the crypto-asset white paper in accordance with Article 8;\n(d) has published the crypto-asset white paper in accordance with Article 9;\n(e) has drafted the marketing communications, if any, in respect of that crypto-asset in accordance with Article 7;\n(f) has published the marketing communications, if any, in respect of that crypto-asset in accordance with Article 9;\n(g) complies with the requirements for offerors laid down in Article 14."},

{"ref": "2", "content": "2. Paragraph 1, points (b), (c), (d) and (f), shall not apply to any of the following offers to the public of crypto-assets other than asset-referenced tokens or e-money tokens:\n(a) an offer to fewer than 150 natural or legal persons per Member State where such persons are acting on their own account;\n(b) ..."},

{"ref": "3", "content": "3. This Title shall not apply to offers to the public of crypto-assets other than asset-referenced tokens or e-money tokens where any of the following apply:\n(a) the crypto-asset is offered for free;\n(b) ..."}]

],

"path": []

```
}  
]  
}
```

BR example

JSON

```
{  
  "regulation": {  
    "title": "Resolução BCB N° 195 de 3/3/2022",  
    "url":  
    "https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=195"  
  },  
  "articles": [  
    {  
      "title": "Art. 47",  
      "content": "Art. 47. As tarifas são cobradas:\nI - do participante  
recebedor, em função da liquidação de ordens de pagamento instantâneo, nos  
termos do art. 40, § 1º; e\nII - do participante solicitante, em função do  
tamanho dos resultados das consultas de lançamentos em Conta PI, de que trata o  
art. 16, inciso III, alínea “c”.",  
      "items": [  
        {"ref": "I", "content": "I - do participante recebedor, em função da  
liquidação de ordens de pagamento instantâneo, nos termos do art. 40, § 1º;  
e"},  
        {"ref": "II", "content": "II - do participante solicitante, em função  
do tamanho dos resultados das consultas de lançamentos em Conta PI, de que  
trata o art. 16, inciso III, alínea “c”."}  
      ],  
      "path": ["CAPÍTULO VIII - DAS TARIFAS"]  
    },  
    {  
      "title": "Art. 48",  
      "content": "Art. 48. A apuração, a cobrança e o pagamento dos valores  
devidos ocorrem no âmbito do ressarcimento de custos do Sisbacen, na forma da  
regulamentação em vigor.\nParágrafo único. A STN, assim como os demais  
participantes diretos, será inserida nos processos de apuração e de rateio do  
ressarcimento dos custos, mas estará dispensada do pagamento de tarifas.",  
      "items": [  

```

```

        {"ref": "", "content": "Art. 48. A apuração, a cobrança e o pagamento dos valores devidos ocorrem no âmbito do ressarcimento de custos do Sisbacen, na forma da regulamentação em vigor."},
        {"ref": "Parágrafo único", "content": "Parágrafo único. A STN, assim como os demais participantes diretos, será inserida nos processos de apuração e de rateio do ressarcimento dos custos, mas estará dispensada do pagamento de tarifas."}
    ],
    "path": ["CAPÍTULO VIII - DAS TARIFAS"]
}
]
}

```

UK example (COBS)

```

JSON
{
  "regulation": {
    "title": "COBS Conduct of Business Sourcebook",
    "url": "https://www.handbook.fca.org.uk/handbook/COBS/1/1.html"
  },
  "articles": [
    {
      "title": "COBS 1.1 General application",
      "content": "COBS 1.1.1 R\nThis sourcebook applies to a firm with respect to the following activities ...",
      "items": [
        {"ref": "COBS 1.1.1 R", "content": "This sourcebook applies to a firm with respect to the following activities carried on from an establishment maintained by it, or its appointed representative, in the United Kingdom:\n(2) designated investment business;\n(3) long-term insurance business in relation to life policies;\nand activities connected with them."},
        {"ref": "COBS 1.1.1A R", "content": "This sourcebook applies to a firm with respect to activities carried on in relation to deposits from an establishment maintained by it, or its appointed representative, in the United Kingdom only as follows: ..."}
      ],
      "path": []
    },
    {

```

```

    "title": "COBS 2.1 Acting honestly, fairly and professionally",
    "content": "COBS 2.1.1\n(1) A firm must act honestly, fairly and professionally in accordance with the best interests of its client (the client's best interests rule). ...",
    "items": [
      {
        "ref": "COBS 2.1.1 R",
        "content": "(1) A firm must act honestly, fairly and professionally in accordance with the best interests of its client (the client's best interests rule).\n(2) This rule applies:\n(a) in relation to designated investment business carried on for a retail client;\n(b) in relation to MiFID, equivalent third country or optional exemption business, for any client; and\n(c) in relation to insurance distribution, for any client.\n(3) For a management company, this rule applies in relation to any UCITS scheme the firm manages."
      },
      {
        "ref": "COBS 2.1.2 R",
        "content": "A firm must not, in any communication relating to designated investment business seek to:\n(1) exclude or restrict; or\n(2) rely on any exclusion or restriction of;\nany duty or liability it may have to a client under the regulatory system."
      }
    ],
    "path": []
  }
]
}

```

AE example (Consumer Protection Standards) (excerpt)

```

JSON
{
  "regulation": {
    "title": "Consumer Protection Standards",
    "url": "https://www.centralbank.ae/media/5crd24gm/cp-standards-pdf.pdf"
  },
  "articles": [
    {
      "title": "1 Definitions",
      "content": "The Definitions set out in Article 1 of the Consumer Protection Regulation (Circular\nNo. 8 – 2020) apply to these Standards",
      "items": [
        {
          "ref": "",
          "content": "The Definitions set out in Article 1 of the Consumer Protection Regulation (Circular No. 8 – 2020) apply to these Standards"
        }
      ],
      "path": []
    }
  ]
}

```



```
    },
    {
      "title": "2.1.1 General Provisions for all Financial Products and/or Services",
      "content": "General Requirements\n2.1.1.1 Licensed Financial Institutions must apply the Disclosure and Transparency requirements to all Financial Products and/or Services provided through all communication channels of service including branches, telephone banking, mobile applications, internet banking and all other channels.\n2.1.1.2 All disclosure information must be available in Arabic and English. Refer to Clause 2.3.1.3.\n2.1.1.3 Disclosure information must be easily available in all branches and all other communication and distribution channels.\n2.1.1.4 Information must be available in a format accessible and in a manner suitable for People of Determination or a representative nominated by such a Consumer. The Licensed Financial Institutions must assess and provide the information in the format best suitable for the Consumer.\n...",
      "items": [
        {"ref": "2.1.1.1", "content": "Licensed Financial Institutions must apply the Disclosure and Transparency requirements to all Financial Products and/or Services provided through all communication channels of service including branches, telephone banking, mobile applications, internet banking and all other channels."},
        {"ref": "2.1.1.2", "content": "All disclosure information must be available in Arabic and English. Refer to Clause 2.3.1.3."}
      ],
      "path": []
    }
  ]
}
```