



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Master's Thesis Nr. 478

Systems Group, Department of Computer Science, ETH Zurich

Brel - A python library for XBRL

by

Robin Schmidiger

Supervised by

Prof. Gustavo Alonso, Dr. Ghislain Fourny

September 2023 – March 2024

DINFK



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

First name(s):

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Signature(s)

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Contents

1	Preface	4
	Preface	4
	Abstract	5
2	Introduction	6
2.1	Introduction	6
2.2	Goals of this thesis	7
2.3	Structure of this thesis	7
3	Related work	9
3.1	XBRL Specification	9
3.2	The XBRL Book	9
3.3	Arelle	9
3.4	Xule	10
3.5	EDGAR	10
3.6	ESEF filings	10
3.7	SEC - Interactive Data Public Test Suite	10
3.8	ESMA Conformance Suite	10
4	XBRL	11
4.1	Overview	11
4.2	Facts	12
4.3	Concepts	13
4.3.1	Taxonomy	13
4.3.2	Concepts	14
4.4	QNames	15
4.5	Segway to advanced XBRL	15
4.6	Networks	16
4.6.1	Types of Networks	16
4.6.2	presentationLink	17
4.6.3	Motivation for Report Elements	18
4.6.4	calculationLink	19
4.6.5	definitionLink	21
4.6.6	labelLink	21
4.6.7	referenceLink	23
4.6.8	footnoteLink	24
4.7	Roles	24
4.8	Hypercubes	25
4.8.1	Dimensions	26
4.8.2	Explicit dimensions	26

4.8.3	Typed dimensions	26
4.8.4	Line items and hypercubes	27
4.9	XBRL Epilogue	27
5	API	29
6	Implementation	31
6.1	Overview	31
6.2	QNames and namespace normalization	31
6.2.1	Overview	31
6.2.2	Namespace hierarchy notation	32
6.2.3	Flattening namespace bindings	33
6.2.4	Collisions in namespace bindings	34
6.2.5	Types of namespace collisions	34
6.3	Implementation of Facts	38
6.3.1	Overview	38
6.3.2	Facts in XBRL	38
6.3.3	Contexts in XBRL	38
6.3.4	Units in XBRL	40
6.3.5	Concepts in XBRL	40
6.3.6	Facts in Brel	40
6.4	Components Implementation	41
6.4.1	Overview	41
6.4.2	Components in XBRL	41
6.4.3	Components in Brel	42
6.5	Implementation of Networks	42
6.5.1	Overview	42
6.5.2	Common Interface	43
6.5.3	Parsing Links into Networks	46
6.5.4	Resolving Locators	46
6.5.5	Parsing Resources	47
6.5.6	Implications of Networks	47
6.5.7	Add Labels to Report Elements	47
6.5.8	LineItems Promoting	47
7	Results	49
8	Conclusion	51
	Bibliography	55

Chapter 1

Preface

This thesis was written as part of my master's degree in computer science at ETH Zurich. It was supervised by Prof. Gustavo Alonso and Dr. Ghislain Fourny. Both the thesis and the source code of the library are available on GitHub[20]. The contents of this thesis are based heavily on both the XBRL standard[7] created by Charles Hoffman and "The XBRL Book"[6] by Ghislain Fourny.

Abstract

TODO: Write the abstract

Chapter 2

Introduction

2.1 Introduction

In the era of data-driven decision making, the ability to efficiently interpret and analyze business reports is becoming increasingly important. Until around 20 years ago, most business reports were published in paper form. This made it difficult for computers to automatically process the information contained in the reports. However, this changed with the introduction of the eXtensible Business Reporting Language (XBRL) in 2000[18]¹. XBRL is a standardized format for business reports that is both machine-readable and can produce human-readable reports. Originally designed for financial reports, XBRL is now used in a wide variety of business reports.[16] Initially, XBRL was based on XML, but it has since been extended to support other formats such as JSON and CSV.

Back in the day, XBRL was a niche technology that was only used by a handful of companies. Now, XBRL is gaining traction in both the public and private sectors. Reporting authorities such as the US Securities and Exchange Commission (SEC) and the European Banking Authority (EBA)[2] are increasingly requiring companies to submit their reports in XBRL format.[26] In the private sector, XBRL is used by companies such as JP Morgan Chase, Microsoft, and Hitachi to automate their financial reporting processes.[3]

However, XBRL is not without its problems. As a standard, XBRL is very complex. Many of its intricacies are a result of legacy design decisions and make working with XBRL unnecessarily difficult. On top of that XBRLs design is closely tied to the XML format, even though standard is not strictly tied to XML anymore.

Since XBRL is a tool primarily designed for non-technical users, ease of use is paramount. However, the current state of XBRL is far from ideal. There are lots of different XBRL tools, most of which are proprietary and closed-source. Some open source tools exist, but apart from the XBRL platform Arelle[19], they are either incomplete or poorly documented.

In 2021, XBRL International published a partial rewrite of the XBRL standard called Open Information Model (OIM)[17]. The goal of OIM is to learn from the mistakes of the past and to make XBRL more accessible to both developers and non-developers. Compared to the original XBRL standard, the OIM is not strictly tied to the XML format. The OIM is a major step in the right direction, but it is still in its infancy and only reworks some of the core XBRL concepts. Also, at the time of writing, the design of the OIM is not yet finalized.

¹Marks the release of XBRL 1.0

2.2 Goals of this thesis

The goal of this thesis is to create a new open source XBRL library that is based on the XBRL, notably the OIM. The name of the library is **Brel** (short for **B**usiness **R**eporting **E**xtensible **L**ibrary) and it is written in the Python programming language. The library should be easy to use and well documented. It should provide a simple python API that allows developers to easily read XBRL reports and extract information from them. Fundamentally, the library should act as a pythonic wrapper around all elements of an XBRL report. The library should also provide basic functionality for validating XBRL reports. Finally, the library should support XBRL reports in XML, but it should be designed in a way that allows it to support other formats in the future.

The research questions that this thesis aims to answer are:

- How can the OIM be translated into an easy-to-use python API?
- How should the non-OIM sections of XBRL be converted into a python API that is consistent with the OIM?
- How can the library be designed to support multiple formats in the future?

2.3 Structure of this thesis

Since this thesis is about XBRL, it is important to first understand what XBRL is and how it works. Therefore, chapter 4 will give a brief introduction to XBRL. It will introduce the core concepts of XBRL in the OIM and move on to the non-OIM sections of XBRL. It will not dive deep into the technical details of the XBRL standard, but instead focus on the concepts that are relevant for this thesis.

The next chapter 5 will introduce the API of Brel. This chapter is already part of the result of this thesis, but it makes more sense to introduce the API before its implementation. The API chapter will answer research question 2.2 and 2.2.

Chapter 6 details the implementation of the Brel library, linking the API described in Chapter 5 with the XBRL standard outlined in Chapter 4. Due to Brel's extensive and complex structure, this chapter will only focus on key aspects of its implementation. While most XBRL to Python translations are straightforward, certain cases are more complex. The chapter will focus primarily on the latter, and explain how Brel handles them. The chapter will also discuss the library's design, specifically its capability to support various formats in future versions, addressing Research Question 2.2.

The second part of the results chapter 7 will evaluate the library based on the goals set out in the introduction. This chapter will also give a comprehensive overview of how brel performs in terms of various XBRL conformance suites. Additionally, it will give examples of how the library can be used to read and validate XBRL reports.

Lastly, chapter 8 will summarize the results of this thesis and give an outlook on future work.

Chapter 3

Related work

The goal of this thesis is to create a new open source XBRL library that is largely based on the OIM. When it comes to work related to Brel, there are three main areas of interest. First, there is the XBRL specification itself and interpretations of it. Second, there are other XBRL libraries and platforms that are similar to Brel. This also includes public databases of XBRL reports. Third, there are the requirements set by authorities and other organizations that XBRL processors have to fulfill.

3.1 XBRL Specification

As mentioned in section 2, this thesis is based on the XBRL standard[7] originally created by Charles Hoffman. The XBRL standard is a complex standard consisting of many different parts. The parts of the XBRL standard that are relevant to this thesis are the Open Information Model (OIM)[17], the XBRL 2.1 specification[9], the extension for dimensional reporting[10] and the specification for generic links[11], all of which will be covered in chapter 4.

The XBRL 2.1 specification and the OIM overlap in most areas, but the OIM is a partial rewrite of the XBRL 2.1 specification. The OIM does not contain all the features of the XBRL 2.1 specification, but it is a lot easier to understand. The OIM also covers some aspects of XBRL dimensions, but does not contain any features from XBRL generic links.

3.2 The XBRL Book

To properly understand the XBRL specification, one already needs to have a good understanding of both XML and XBRL. This makes it difficult for beginners to get started with XBRL. To address this problem, Dr. Ghislain Fourny wrote "The XBRL Book"[6]. The book is a comprehensive guide for XBRL and covers all the important aspects of the XBRL standard, including the rather recent OIM.

3.3 Arelle

ArELLE[19] is an open source XBRL platform. At the time of writing, Arelle is the most fully open source platform of its kind. It supports all the features of the XBRL 2.1 specification and the OIM. Like Brel, Arelle is also written in Python and is open source. Unlike Brel, Arelle is a complete XBRL platform and not just a python library. Think of Arelle as "Excel for XBRL".

3.4 Xule

Xule[28] is a rule language for XBRL. It is a declarative language that allows users to write rules for validating XBRL reports. It is part of the Arelle project and is used by Arelle to validate reports. Think of Xule as a domain-specific language for XBRL validation rules. Even though Xule is not part of this thesis, Brel should be able to support Xule in the future.

3.5 EDGAR

The SEC (US Securities and Exchange Commission) maintains a system called EDGAR [24].¹ EDGAR is a database of financial reports submitted to the SEC. EDGAR stands for **E**lectronic **D**ata **G**athering, **A**nalysis, and **R**etrieval. Filings submitted to EDGAR have to be in XBRL format. The SEC provides public access to the filings submitted to EDGAR.²

3.6 ESEF filings

The European counterpart to the SEC is called ESMA (**E**uropean **S**ecurities and **M**arkets **A**uthority). It also maintains a database of ESEF filings[21]. ESEF stands for **E**uropean **S**ingle **E**lectronic **F**ormat, which is a standard for XBRL reports in the EU. Similar to EDGAR, the ESEF database is publicly accessible.³ An interesting aspect of this database is that it is hosted by XBRL International, the organization that maintains the XBRL standard.

3.7 SEC - Interactive Data Public Test Suite

To ensure that the XBRL processors that the companies use to create their XBRL reports are compliant with the XBRL standard, the SEC created the Interactive Data Public Test Suite[23]. This test suite contains an extensive collection of XBRL reports that are used to test XBRL processors. The SEC provides this test suite free of charge and it is available on their website.

Brel uses this test suite to test its XBRL processor. At the time of writing, Brel does not pass all the tests in the test suite, since Brel does not support all the features of the XBRL standard. However, building a processor that passes all the tests is not a feat that can feasibly be accomplished in the scope of this thesis.

3.8 ESMA Conformance Suite

The European Securities and Markets Authority (ESMA) also maintains a test suite for XBRL processors[22]. This test suite is similar to the SEC's Interactive Data Public Test Suite, but it is maintained by a different authority. Another key difference is that the ESMA Conformance Suite facilitates automated testing of xHTML reports.

Brel does not support xHTML reports, so the ESMA Conformance Suite is not relevant for this thesis. However, Brel should be able to support xHTML reports in the future.

¹not to be confused with the U.S. SEC, which stands for U.S. Soybean Export Council

²<https://www.sec.gov/edgar/search/>

³<https://filings.xbrl.org/>

Chapter 4

XBRL

4.1 Overview

The content of this thesis is largely based on the XBRL standard[7] created by Charles Hoffman and Dr. Ghislain Fourny's interpretation of it in *the XBRL Book*[6]. Since the thesis builds on the foundation laid by the two, it is important to understand the ground work that they have done. This chapter will give a brief introduction to XBRL. Think of it as a crash course in XBRL.

In essence, XBRL is a standardized format for representing reports. After all, XBRL stands for **eXtensible Business Reporting Language**. [7]

As Ghislain Fourny has put it in *the XBRL Book* [6]:

If XBRL could be summarized in one single definition, it would be this:
XBRL is about reporting facts.

Keeping this in mind, the subsequent sections will first introduce the basic concepts of XBRL, namely facts, concepts and QNames. Afterwards, I will introduce the more advanced concepts that are about putting facts into relation with each other, namely roles, networks, and report elements.

Armed with this fundamental knowledge about XBRL, you will then be able to understand how Brel implements the core parts of the standard and how it hides a lot of the complexity of XBRL behind a simple Python API.


This chapter will not cover the XBRL specification in its entirety. It will also gloss over a lot of the details of the specification. It is more focused on giving the reader a high-level overview of the contents of the XBRL specification.

Furthermore, a lot of concepts in XBRL require knowledge of other XBRL concepts. There are also a few circular dependencies between the concepts, which makes it hard to explain them in a linear fashion. Therefore, there are a few sections in this chapter that will redefine concepts that have already been introduced. This is done to gradually introduce the reader to the more complex concepts of XBRL.

4.2 Facts

A fact is the smallest unit of information in an XBRL report. The word "Fact" is a term used to describe an individual piece of financial or business information within an XBRL instance document.

Lets consider a simplified example involving a financial report of Microsoft Corporation for the fiscal year 2022. Microsoft's annual report can be found on the company's website⁰. It contains a lot of information about the company's financial situation, as well as information about the company's business activities. For this example, we will only consider the company's revenue for the fiscal year 2022. Microsoft chose to report this information as follows:



Annual Report
2022

SUMMARY RESULTS OF OPERATIONS

(In millions, except percentages and per share amounts)	2022	2021
Revenue	\$ 198,270	\$ 168,088
Gross margin	135,620	115,856
Operating income	83,383	69,916
Net income	72,738	61,271
Diluted earnings per share	9.65	8.05
Adjusted net income (non-GAAP)	69,447	60,651
Adjusted diluted earnings per share (non-GAAP)	9.21	7.97

Figure 4.1: Microsoft's summary results of operations for the fiscal year 2021 and 2022

[5]

This table contains multiple facts about Microsoft for both fiscal years 2021 and 2022, as seen by the horizontal axis. The vertical axis describes what is being reported. The "what is being reported" part is called the **concept** of a fact. It reports the values for the concepts "Revenue", "Gross margin", "Operating income", ... for both fiscal years. In summary, the table contains 14 facts across 7 concepts for 2 fiscal years.

For now, let us focus on the top left fact, which reports the company's revenue for the fiscal year 2022. In XBRL, a corresponding fact would be represented as follows:

- **Concept:** Revenue
- **Entity:** Microsoft Corporation
- **Period:** from 2022-04-01 to 2023-03-31⁰
- **Unit:** USD
- **Value:** 198'270'000¹

In this example:

- The **concept** refers *what* is being reported. In this case, "Revenue" indicates that the fact is reporting information about the company's revenue.

⁰<https://www.microsoft.com/investor/reports/ar22/index.html>

⁰Refers to the fiscal year 2022, which starts on April 1, 2022 and ends on March 31, 2023

¹<https://www.microsoft.com/investor/reports/ar22/index.html>

- The **entity** refers to *who* is reporting. In the case of our example, the entity is "Microsoft Corporation". In our example this is implicit, since we are looking at Microsoft's annual report. However, the entity of a fact has to be explicitly stated in an XBRL report.
- The **period** refers to *when* the information is being reported. The period is defined as the fiscal year 2022. This is indicated by the column header "2022" in the table.
- The **unit** refers to *how* the information is being reported. In this example, the unit is "USD", which indicates that the information is being reported in US dollars. The unit is indicated by the dollar sign \$ in the table.
- The **value** refers to *how much* is being reported. According to Microsoft's 2022 annual report, the company's revenue for the fiscal year 2022 was around 198 billion US dollars.

The concept, entity, period and unit of a fact are called its **aspects**. If necessary, additional aspects can be defined for a fact. These additional aspects are called **dimensions** and will be covered in section 4.8 The aspects that are not dimensions are called **core aspects**. Even though the name suggests otherwise, core aspects are not all mandatory for a fact. The only mandatory core aspect is the concept.

4.3 Concepts

In section 4.2, we learned that a fact is the smallest unit of information in an XBRL report. One of the core aspects of a fact is its concept, which refers to what is being reported.

So for example, if a fact is reporting information about a company's revenue, then the concept of the fact is "Revenue". In this section, we will take a closer look at concepts and how they are defined in XBRL.

Concepts are the fundamental building blocks of XBRL and they are defined in what is called the **taxonomy**.

4.3.1 Taxonomy

Simply put, the XBRL taxonomy is a collection of concepts. Each XBRL report defines its own taxonomy inside of a taxonomy schema file. The taxonomy defined by the report is called the **extension taxonomy**.

This extension taxonomy contains references to other taxonomies, which may contain references to even more taxonomies. So when a report and its extension taxonomy are loaded into memory, the entire taxonomy is loaded into memory. The transitive closure of all these references is called the **DTS** (short for **D**iscoverable **T**axonomy **S**et).

Note that most of the taxonomies in the DTS are not located on the same machine as the report. Instead, they are located on the internet and are downloaded on demand.

Some taxonomies commonly found in a report's DTS are:

- **us-gaap**¹ - Contains concepts for US Generally Accepted Accounting Principles (GAAP).
- **ifrs**² - Contains concepts for International Financial Reporting Standards (IFRS).

¹<https://xbrl.us/us-gaap/>

²<https://www.ifrs.org/>

- **dei**³ - Contains concepts for the SEC's Document and Entity Information (DEI) requirements.
- **country**⁴ - Contains concepts for country codes.
- **iso4217**⁵ - Contains concepts for currency codes.

Since a lot of DTSs from different reports share a lot of the same taxonomies, it makes sense to cache the taxonomies locally, instead of downloading them every time they are needed.

4.3.2 Concepts

Each concept in the DTS is identified by a **QName**. The technical intricacies of QNames will be covered in section 4.4, but for now think of them as a unique identifier for a concept. The QName of a concept tends to be human-readable and self-explanatory. However, accountants and business analysts tend to go overboard with their naming conventions. Some examples of the QNames of concepts are:

- **us-gaap:Assets**
- **ko:IncrementalTaxAndInterestLiability**
- **dei:EntityCommonStockSharesOutstanding**
- **us-gaap:ElementNameAndStandardLabelInMaturityNumericLowerEndTo-NumericHigherEndDateMeasureMemberOrMaturityGreaterThanLowEnd-NumericValueDateMeasureMemberOrMaturityLessThanHighEndNumeric-ValueDateMeasureMemberFormatsGuidance**

But concepts do not only consist of a QName. They also constrain some of the aspects and values of the facts that reference them. Going back to our running example from section 4.2, the concept **us-gaap:Revenue** introduces some constraints. For example, it constrains the value to be a **monetaryItemType**. This means that the value of the fact must be a number, not just any arbitrary string.⁶ It also constrains the unit to be a currency defined in the ISO 4217 standard.[1] Moreover, monetary facts have to be labeled as either "debit" or "credit" using the **balance** attribute. In this case, the concept **us-gaap:Revenue** constrains the fact to have a "debit" balance, since the company's revenue is an asset.

The concept **us-gaap:Revenue** also constrains the period of the fact to be of type "duration". This means that the period of the fact has to be a duration of time, such as a fiscal year or a quarter. Alternatively, the period could be of type "instant", which means that the period refers to a specific point in time.

Finally, the concept **us-gaap:Revenue** allows the fact to be null, which means that it is optional for a report to contain a fact for the concept **us-gaap:Revenue**. The vast majority of concepts allow facts to be null.

This wraps up our discussion about concepts in XBRL. In the next section, we will take a look at QNames and how they are used in XBRL. Together with the knowledge about concepts and facts, this will give us a solid foundation to understand the core parts of the XBRL standard.

³<https://www.sec.gov/info/edgar/dei-2019xbml-taxonomy>

⁴<https://xbrl.fasb.org/us-gaap/2021/elts/us-gaap-country-2021-01-31.xsd>

⁵<https://www.iso.org/iso-4217-currency-codes.html>

⁶The **monetaryItemType** has additional constraints besides being an integer, but we will ignore them for now.

4.4 QNames

Although the motivation behind the XBRL processor Brel is to shield its user from the complexity of XML, we keep one key aspect of XML in our API: QNames.

QNames are a way to uniquely identify an XML element or attribute. They consist of three things: a namespace prefix, a namespace URI, and a local name. The prefix acts as a shorthand for the namespace URI.

For example the QName `us-gaap:Assets` identifies the element `Assets` in the namespace `us-gaap`.

In this example, the namespace prefix `us-gaap` is a shorthand for the namespace URI `https://xbrl.fasb.org/us-gaap/2022/elts/us-gaap-2022.xsd`, and together they form the namespace `us-gaap`.

Figure 4.2: The `us-gaap:Assets` QName

- Namespace prefix: `us-gaap`
- Namespace URI: `https://xbrl.fasb.org/us-gaap/2022/elts/us-gaap-2022.xsd`
- Local name: `Assets`

QNames are used in XBRL to identify concepts, facts and other elements. Since they provide a robust and easy way for identifying elements, we decided to use them in our API as well. However, there is one important difference between our QNames and the QNames used in the XBRL taxonomy: Currently, most XBRL filings are based on XML, where namespace prefixes are defined on a per-element basis. Therefore, the mapping from namespace prefixes to namespace URIs depends on where the QName is used.

In our API, there is a fixed, global mapping from namespace prefixes to namespace URIs. The motivation behind this decision is that it makes the API easier to use. More details about this mapping will be explained in section 6.2.

4.5 Segway to advanced XBRL

This concludes our overview of QNames as well as the overview of the core concepts of XBRL. Armed with this knowledge, we could already create functional XBRL reports, albeit with a lot of limitations.

The most glaring limitation is that the facts in the report are not structured in any way. Reports are just an unstructured set of facts. Since facts are not related to each other, it is impossible to verify if the values within the report are consistent. Besides that, XBRL in its current form is not very user-friendly. The concepts for example are named using QNames, which tend to be human-readable but extremely verbose. Another limitation is that QNames are mostly in English, which makes it difficult for non-English speakers to understand the report.

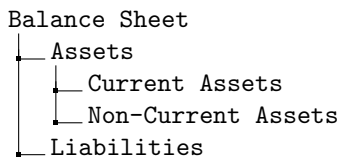
All of these issues will be addressed in the next sections, which will introduce the advanced concepts of XBRL, the most important of which are networks.

4.6 Networks

Up to this point, all concepts in the XBRL taxonomy are treated as independent entities. The whole filing can be viewed as an unordered set of facts with no relation between them. However, in reality, concepts are not independent of each other. Instead, they are related to each other in some way.

For example, the concepts **Assets** and **Liabilities** are related to each other in the sense that they are both part of the concept **Balance Sheet**. Furthermore, the concept **Assets** can be further divided into the concepts **Current Assets** and **Non-Current Assets**. The aforementioned relations can be visualized using a directed graph, as shown in figure 4.3.

Figure 4.3: Example of a relations between concepts in XBRL



A reader with a basic understanding of mathematics will recognize that the above example is a directed acyclic graph (DAG). More specifically, the above example is a tree. Graphs are commonly used to represent relations between entities. In the context of XBRL, graphs are used to represent all kinds of relations between concepts, facts, and other elements of an XBRL filing. From now on, I will refer to graphs that represent relations between XBRL elements as **networks**.

XBRL commonly uses the term **Extended Link** or **Link** to refer to networks or parts of networks. I will use the term **network** throughout this thesis, since it is more intuitive and less ambiguous. When I refer to a **link**, I am referring to something specific in the XBRL specification. When I refer to a **network**, I am referring to the general concept of a network.

XBRL groups multiple links together into so called **linkbases**. From a semantic perspective, linkbases do not have any meaning. From a technical perspective, linkbases are just XML elements that have children that are links.

4.6.1 Types of Networks

The XBRL 2.1 specification defines 6 built in types of networks[15]¹:

- **link:presentationLink**: A network that represents the hierarchy of concepts in a report. An example of this can be seen in figure 4.3.
- **link:calculationLink**: A network that represents how concepts are calculated from other concepts. For example, in figure 4.3, the concept **Assets** is calculated as the sum of the concepts "Current Assets" and "Non-Current Assets".
- **link:definitionLink**: A network that represents the typing relations between concepts. The meaning of this network will become clearer in section 4.6.5.
- **link:labelLink**: A network that associates report elements with human-readable labels.

¹The XBRL 2.1 specification is inconsistent about **link:footNotelink**. Section 1.4 does not list it as a standard extended link, section 3.5.2.4 does. I will assume that it is a standard extended link.

- **link:referenceLink:** A network that links report elements to external resources. For example, the concept `Total Shareholder Return Amount` might have an official definition in the SEC's Code of Federal Regulations (CFR). The reference network would link the concept to the subparagraph 17 CFR 229.402(v)(2)(iv)[27].
- **link:footnoteLink:** A network that links report elements, facts and other elements to footnotes.

XBRL refers to these built in networks as **standard extended links**. If needed, XBRL allows users to define their own networks, which are referred to as **custom extended links**[15].

Technically speaking, XBRL does allow networks in XBRL to contain both directed and undirected cycles. However, in practice, networks in XBRL are almost always directed acyclic graphs (DAGs).

In the subsequent sections, I will describe how networks are implemented in XBRL on a conceptual level.

Even though labels will be covered in more detail in section 4.6.6, I will already use them throughout this chapter. Labels are used to associate report elements with human-readable labels. For example, the concept `us-gaap:CurrentAssets` might have the label "Current Assets" in the English language. The main motivation behind this editorial decision is that it makes the chapter easier to read.

4.6.2 presentationLink

The **link:presentationLink** network is used to represent the hierarchy of concepts in a report. I will describe presentationLinks in more detail compared to the other networks, since all other network types are implemented in a similar fashion as presentationLinks.

XBRL implements all its networks as a list of directed edges called **arcs**. Each arc has a source and a target. Duplicate arcs are not allowed.

Taking the example from figure 4.3, the presentationLink network would be represented as the following edge list:

Figure 4.4: Example of a presentationLink network in edge list format

```
Balance Sheet -> Assets
Assets -> Current Assets
Assets -> Non-Current Assets
Balance Sheet -> Liabilities
```

Each arc in the example 4.4 is represented as a **link:presentationArc** element in XBRL. Besides presentationArcs, presentationLinks contain so called "locators" **link:loc** that represent the nodes in the network. In case of a presentation network, the locators are references to the concepts in the XBRL taxonomy. In other networks, locators can be references to other elements, such as facts.

Going back to the example in figure 4.4, the first arc `Balance Sheet -> Assets` would be represented as follows in XML syntax:

```

<link:loc
  xlink:type="locator"
  xlink:href="file_1.xsd#BalanceSheet"
  xlink:label="BalanceSheet_loc"
/>
<link:loc
  xlink:type="locator"
  xlink:href="file_1.xsd#Assets"
  xlink:label="Assets_loc"
/>
<link:presentationArc
  xlink:type="arc"
  xlink:arcrole="http://www.xbrl.org/2003/arcrole/parent-child"
  xlink:from="BalanceSheet_loc"
  xlink:to="Assets_loc"
  order="1"
/>

```

Figure 4.5: Example of a presentationArc in XML syntax

The XML snippet in figure 4.5 contains two locators and one arc. The two locators represent the nodes of the arc, referencing the concepts **BalanceSheet** and **Assets** respectively. The arc represents the edge between the two nodes.

Let us look at the XML snippet 4.5 step by step.

- **Specify types:** The `xlink:type` attribute specifies the types for both the locators and the arc. For the former, the type is `locator`, whereas for the latter, the type is `arc`.
- **Connect nodes and edges:** Both locators contain an `xlink:label` attribute that uniquely identifies the locator. The arc links the two locators together using the `xlink:from` and `xlink:to` attributes.
- **Order of edges:** The outgoing edges of a node are ordered using the `order` attribute of the arc.
- **Specify arcrole:** The `xlink:arcrole` attribute of the arc specifies the kind of relation between the source and the target of the arc. In case of a presentationLink, the `xlink:arcrole` attribute is always set to `parent-child`.

Locators and arcs form the basic building blocks of all networks in XBRL, notably presentationLinks. A presentationLink is just a container for locators and arcs. To build a fully featured presentation network such as in figure 4.3, we need to add more locators and arcs to the presentationLink. To avoid cluttering the chapter with XML snippets, I will only describe the XML syntax for the first arc in the network.

4.6.3 Motivation for Report Elements

Up to this point, I have only described how presentation networks are implemented in XBRL. I also mentioned that presentation networks introduce a hierarchy of concepts, but this is not entirely true.

I have introduced concepts as the "what"-part of a fact. For example, if the company Foo reports a revenue of 1000 USD in 2019, the concept **Revenue** is the "what"-part of the fact.

However, if we look at the presentation network in figure 4.3, not all of the elements in the network can have a fact associated with them. An example of this is the concept `BalanceSheet`. In XBRL, concepts that can not have a fact associated with them are called `Abstract`.

The Open Information Model (OIM) combines abstracts and concepts using the term `report element`. Report elements are, as the name suggests, elements that can appear in a report. Some of these report elements are used for facts, namely the concepts. Others are used to represent the structure of the report, namely the abstracts. There are six types of report elements in total[17]. I will introduce them as they come up in the subsequent sections.

With the introduction of report elements, our notion of a presentation network changes slightly. Instead of introducing a hierarchy of concepts, presentation networks introduce a hierarchy of report elements. However, our notion of a fact stays the same. A fact is still requires a concept, not a report element.

4.6.4 calculationLink

The `link:calculationLink` network is used to represent how concepts are calculated from other concepts. More specifically, it is used to represent how a concept is the sum of other concepts. Under the hood, calculationLinks are implemented in the same way as presentationLinks, but there are a few differences in the semantics.

1. The arcs now have the tag `link:calculationArc` instead of `link:presentationArc`.
2. The link now has the tag `link:calculationLink` instead of `link:presentationLink`.
3. The `xlink:arcrole` attribute of the `link:calculationArc` element is set to `summation-item`.
4. The `link:calculationArc` element has an additional attribute called `weight`.
5. All locators in the link are references to concepts, not just report elements.

Most of these differences are self-explanatory and do not have any semantic implications. However, the last two differences are worth explaining in more detail.

The main motivation behind calculation networks is so that XBRL processors can either calculate the value of a concept or check if the value of a concept is computed correctly and consistently. In the case of our XBRL processor Brel, the main focus is on the latter. Chapter 6.4 of "The XBRL Book" [6] describes the different consistency checks in more detail.

Facts that have a concept within a calculation network are computed as a weighted sum of their children. The `weight` attribute of the `link:calculationArc` element specifies the weight of the child in the sum. Additionally, facts can only be associated with concepts, not just any report element. Therefore, all locators in a calculation network are references to concepts.

In the section on concepts4.3, I have introduced the `balance` aspect of a concept. It specifies if the concept is a debit or a credit. The XBRL 2.1 specification enforces some constraints on the `balance` aspect of concepts in combination with the `weight` attribute of the `link:calculationArc` element [12]. If one concept has a `balance` of `debit` and another concept has a `balance` of `credit`, then their connecting arc must have a negative `weight`. If both concepts have the same `balance`, then their connecting arc must have a positive `weight`.

Concept 1	Concept 2	Connecting edge weight
Debit	Credit	≤ 0
Credit	Debit	≤ 0
Debit	Debit	≥ 0
Credit	Credit	≥ 0

Figure 4.6: Balance and weight constraints in calculation networks

A network that is consistent with the balance and weight constraints is called a **balance consistent network**.^[6]

Balance consistency is not the only kind of consistency that calculation networks can be checked for. Another kind of consistency is **roll-up consistency** which comes in two flavors: **simple roll-up consistency** and **nested roll-up consistency**. Simple roll-up is roll-up consistency without any nested concepts. So the calculation network can only have a depth of 1.

Nested roll-up consistency is roll-up consistency with nested concepts. So the calculation network can have a depth of more than 1.

Roll-up consistency requires a calculation network as well as a presentation network and checks if the structure of the two networks is consistent. For example, if the calculation network contains the arc **Assets** \rightarrow **Savings Accounts**, then the presentation network must also contain the arc **Assets** [Abstract] \rightarrow **Savings Accounts**.

Remember that calculation networks can only contain concepts, not abstracts. So the report element **Assets** in the calculation network and the report element **Assets** [Abstract] in the presentation network are not the same. But how does XBRL know that they are related?

The answer is that the two report elements are related by the presentation network. It contains the arc **Assets** [Abstract] \rightarrow **Assets**.

Let us visualize this example in figure 4.7, which shows the calculation network and the presentation network side by side. The two networks are roll-up consistent with each other. The example is expanded a bit and also contains the concepts **UBS Savings Account**, **Raiffeisen Savings Account** and **Liabilities**. Note that the calculation network contains weights for the arcs, but the presentation network does not.

Figure 4.7: Example of nested roll-up consistency

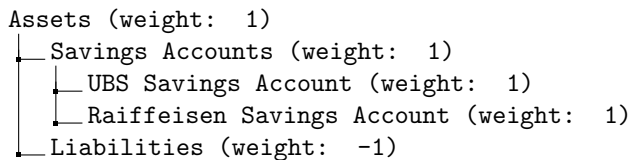


Figure 4.8: Calculation network

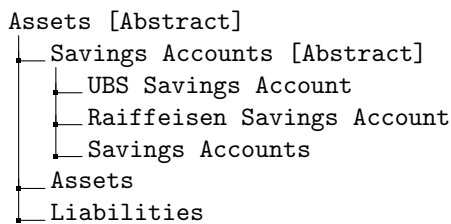


Figure 4.9: Presentation network

The presentation network in figure 4.7 is roll-up consistent with the calculation network. It contains the two abstracts **Assets** [Abstract] and **Savings Accounts** [Abstract], which are not present in the calculation network. They take the role of the concepts **Assets** and **Savings Accounts** in the calculation network.

Besides roll-up and balance consistency, there is a third kind of consistency. XBRL does not have a name for it, but I will refer to it as **aggregation consistency**. It checks if in a calculation network, for each concept, the child concepts add up to their parent concept. Going back to our example in figure 4.7, the values of the UBS Savings Account and the Raiffeisen Savings Account should add up to the value of the Savings Accounts concept. Aggregation consistency uses weighted sums to add up the values of the child concepts, where the weight of a child concept is the weight of the connecting arc.

To check for aggregation consistency, we need to know the values of the facts that are associated with the concepts. So aggregation consistency checks a list of facts against a calculation network.

Aggregation consistency is best explained using an example. Consider the following list of facts 4.10 for the calculation network in figure 4.7:

Figure 4.10: Alice’s savings accounts and liabilities in CHF

Concept	2022	2023
UBS Savings Account	1000	1000
Raiffeisen Savings Account	2000	3000
Savings Accounts	3000	4000
Liability	500	500
Assets	2500	3499

As you can see, there are two facts reported against each concept, one for 2022 and one for 2023.

If there are multiple facts reported against a concept, then we iterate over them and “pin” all aspects except for the concept. Next, we go to the list of all facts and filter out all facts that have the same pinned aspects. Finally, we check for aggregation consistency using the filtered list of facts and the calculation network.

In our example, we would first check all facts for 2022, then all facts for 2023. The list of facts is aggregation consistent for the year 2022, since the values of the UBS Savings Account and the Raiffeisen Savings Account add up to the value of the Savings Accounts concept. Also, the value of the Savings Accounts minus the value of the Liabilities concept adds up to the value of the Assets concept.

However, the list of facts is not aggregation consistent for the year 2023. The reason being that the value of the Savings Accounts minus the value of the Liabilities concept does not result in the value of the Assets concept.

4.6.5 definitionLink

TODO

4.6.6 labelLink

The `link:labelLink` network is used to associate report elements with human-readable labels. Thus far, we referred to report elements using their QName. Obviously, if we were to open an XBRL report in a viewer such as Arelle, we would not be greeted with QNames for the concepts, abstracts, etc. Instead, we would see nicely formatted human-readable labels.

The following figure shows an example of the condensed consolidated statement of income of the Coca Cola Company[4] of Q2 2019 as displayed in Arelle[19].

Concept	2023-06-30
0000002 - Statement - CONDENSED CONSOLIDATED STATEMENTS OF INCOME	
Income Statement [Abstract]	
Statement [Table]	
Statement	
Net Operating Revenues	22,952,000,000
Cost of goods sold	9,229,000,000
Gross Profit	13,723,000,000
Selling, general and administrative expenses	6,506,000,000
Other operating charges	1,338,000,000
Operating Income	5,768,000,000
Interest income	392,000,000
Interest expense	374,000,000
Equity income (loss) — net	813,000,000
Other income (loss) — net	91,000,000
Income Before Income Taxes	2,880,000,000
Income taxes	359,000,000
Consolidated Net Income	5,634,000,000
Net Income (Loss) Attributable to Noncontrolling Interest	-20,000,000
Net Income Attributable to Shareowners of The Coca-Cola Company	5,654,000,000
Basic Net Income Per Share ¹	1.31
Diluted Net Income Per Share ¹	1.30
Average Shares Outstanding — Basic	4,325,000,000
Effect of dilutive securities	18,000,000
Average Shares Outstanding — Diluted	4,343,000,000

Figure 4.11: Statement of income of the Coca Cola Company of Q2 2019

Let us take a closer look at the first concept of the statement of income: Revenue. Even though the concept for the revenue is `us-gaap:Revenues`, Arelle displays it as "Net Operating Revenues".

Arelle achieves this by using the `link:labelLink` network that is part of the XBRL report. LabelLinks are another type of extended link that associates report elements with strings. They are implemented in the same way as presentationLinks and calculationLinks, but this time they not only contain arcs and locators, but also `link:label` elements.

From a semantic perspective, labels are different from report elements such as concepts and abstracts. Instead, they are a kind of **resource**. Resources are essentially metadata about report elements, facts, and other elements of an XBRL report.

The approach that XBRL takes to labels and other resources is quite interesting. Going back to our example in figure 4.11, the definition of the concept `us-gaap:Revenues` happens independently from any labels that are associated with it. The labels are later associated with the concept using the `link:labelLink` network. In fact, each report element can potentially have multiple labels in different languages or different degrees of verbosity. To categorize labels, XBRL uses the concept of **roles**, which are covered in more detail later in this chapter. We have already seen the `xlink:arcrole` attribute in the `link:presentationArc` 4.6.2. The role of a label works in a similar way.

In terms of the XML syntax, the `link:labelLink` network is implemented in the same way as the `link:presentationLink` network. The only addition is the `link:label` element, which is used to represent a label. It contains a few pieces of information:

1. The `xlink:label` attribute, which is used to reference the arc that the label is associated with. This must not be confused with the text of the label. Think of it as a unique identifier for the label.

2. The `xlink:role` attribute, which specifies the role of the label. More on this later.
3. The `xml:lang` attribute, which specifies the language of the label.
4. The `xlink:type` which is always set to `resource`.
5. The actual label text. This is the human-readable label that Arelle displayed in figure 4.11.

For example, for the label "Net Operating Revenues" 4.11, the following XML segment would be used:

```
<link:label
  id="1234"
  xlink:label="lab-us-gaap-Revenues"
  xlink:role="http://www.xbrl.org/2003/role/terseLabel"
  xlink:type="resource"
  xml:lang="en-US">
  Net Operating Revenues
</link:label>
```

Figure 4.12: Example of a label in XML syntax

Note that we have omitted both the connecting arc and the locator in this example, as they work in the same way as in all other networks.

The label in figure 4.12 has the role `http://www.xbrl.org/2003/role/terseLabel`. This role is used to indicate that the label text is short and concise. XBRL defines a few other roles of which the most important ones are:

Role	Description
<code>http://www.xbrl.org/2003/role/label</code>	The default label role.
<code>http://www.xbrl.org/2003/role/terseLabel</code>	A short, human-readable label.
<code>http://www.xbrl.org/2003/role/verboseLabel</code>	A long, human-readable label.
<code>http://www.xbrl.org/2003/role/positiveLabel</code>	A label for positive values.
<code>http://www.xbrl.org/2003/role/negativeLabel</code>	A label for negative values.
<code>http://www.xbrl.org/2003/role/zeroLabel</code>	A label for zero values.
<code>http://www.xbrl.org/2003/role/documentation</code>	A label for documentation.

Figure 4.13: Important label roles

Note that the 4.13 is not exhaustive. There are many more label roles that are used in practice. Users can even define their own label roles. For a complete list of standard label roles, refer to the XBRL 2.1 specification[14].

4.6.7 referenceLink

The `link:referenceLink` network is used to link report elements to external resources. For example, the concept `us-gaap:Revenues` might have an official definition in the SEC's Code of Federal Regulations (CFR). The reference establishes a link between the concept and external resource such as the CFR. Intuitively, think of the reference as a citation in a scientific paper.

Structurally, referenceLinks are implemented in the same way as LabelLinks - they contain arcs, locators, and resources. The only difference is that the resources are

references to external resources, not labels. Whereas labels are mostly just text, references are take the form of dictionaries.

The following figure shows an example of a reference in the XML syntax. Both the accompanying arc and locator are omitted for brevity. The only noteworthy changes to the arc are the tag `link:referenceArc` and the `xlink:arcrole` attribute, which is set to `concept-reference`.

```
<link:reference
  xlink:type="resource"
  xlink:label="SECRegulationS-K229402v2vi"
  xlink:role="http://www.xbrl.org/2003/role/presentationRef"
>
  <ref:Publisher>SEC</ref:Publisher>
  <ref:Name>Regulation S-K</ref:Name>
  <ref:Number>229</ref:Number>
  <ref:Section>402</ref:Section>
  <ref:Subsection>v</ref:Subsection>
  <ref:Paragraph>2</ref:Paragraph>
  <ref:Subparagraph>vi</ref:Subparagraph>
</link:reference>
```

Figure 4.14: Example of a reference for the concept `edc:CoSelectedMeasureName`

The children of the `link:reference` element form a dictionary that describes the external resource that the reference points to.

In the example 4.14, the reference points 17 CFR 229.402(v)(2)(vi) [25]⁷. References can point to any kind of external resource, not just the CFR. They can point to other XBRL reports, PDFs, websites, etc.

Usually, an XBRL report contains only one `referenceLink`, which is used to link the concepts in the report to the concepts to the underlying code of regulations.

4.6.8 footnoteLink

The `link:footnoteLink`, just like the `link:referenceLink` and the `link:labelLink`, is used to associate report elements with resources. One difference is that the resources are footnotes, not labels or references, which is a surface level difference. The other difference is that the locators in the `footnoteLink` can also reference facts, not just report elements. Other than that, `footnoteLinks` are implemented in the same way as the other networks.

4.7 Roles

Even though the networks introduced in the previous section 4.6 provide a good foundation for structuring XBRL reports, they are not sufficient to create a comprehensive overview of the report whole report, only individual sections of it. Moreover, the roll-up consistency of calculation networks 4.6.4 introduced the notion of having networks related to each other. With our current understanding of XBRL, there is no way to express this relationship. This is where **Roles** come into play.

Roles are a way to group networks together into a what is essentially a chapter of a report. Each set of networks is assigned a unique URI and potentially a label.

⁷CFR stands for Code of Federal Regulations.

For example, a report might have a role for the cover page, one for the balance sheet, one for the income statement, and so on. The balance sheet would only contain a presentation network, while the income statement would contain a presentation network, a calculation network, and potentially a definition network.

A role usually contains a presentation network, a calculation network, and a definition network. The other types of networks are not commonly used in roles. Rather, they belong to the report as a whole. An example of this would be a label network that contains all the labels for the entire report.

Roles in the XBRL XML syntax follow a simple structure, which I will explain using an example of a balance sheet role.

Figure 4.15: Example of the role "Balance Sheet" expressed in XBRL XML syntax

```
<link:roleType id="BalanceSheet" roleURI="http://www.foocompany.com/role/BalanceSheet">
  <link:definition>Foo balance</link:definition>
  <link:usedOn>link:presentationLink</link:usedOn>
  <link:usedOn>link:calculationLink</link:usedOn>
</link:roleType>
```

The role in figure 4.15 has the following properties:

- **roleURI** (required): The URI of the role. This URI is used to reference the role from other elements in the XBRL taxonomy. It is the primary identifier of the role.
- **definition** (optional): A human-readable description of the role.
- **usedOn**: A list of links that the role can be used in.

The networks that are associated with the role are not defined in the role itself. Rather, each link that uses the role has to declare the role in the **role** property, which is used to reference the role from the link.

Whenever a link references a role, the role must have a **usedOn** property that contains the type of the link. Going back to figure 4.15, if a definition network would reference the balance sheet role, a conformant XBRL processor would throw an error. This is because the balance sheet role does not declare the **definitionLink** type in its **usedOn** property.

4.8 Hypercubes

One key observation that can be made when looking at the facts of an XBRL report is that they are often structured like a hypercube. The aspects of a fact can be seen as the dimensions of a hypercube, whereas the value of the fact is the value of the hypercube at the given dimensions.

Figure 4.16: Example of a hypercube

Period	Entity	Concept	Value
2020	Foo	Sales	100\$
2020	Foo	Costs	50\$
2020	Bar	Sales	200\$
2020	Bar	Costs	100\$
2021	Foo	Sales	150\$
2021	Foo	Costs	75\$
2021	Bar	Sales	250\$
2021	Bar	Costs	125\$

TODO: 3D image of hypercube

Hypercubes are a common way to structure data nowadays. Yet, back when XBRL was created, they were not as prevalent as they are today. In fact, the early versions of XBRL did not support hypercubes at all. They were retrofitted into the XBRL specification in 2006.[10].

4.8.1 Dimensions

When viewing facts as hypercubes, the cube ends up having four built in dimensions. These correspond to the four core aspects of a fact: **Period**, **Entity**, **Concept**, and **Unit**. XBRL allows for the creation of custom dimensions, which come in two flavors: explicit and typed.

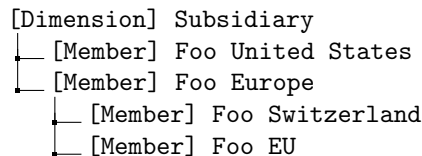
4.8.2 Explicit dimensions

Explicit dimensions are dimensions that have a predefined set of possible values. For example, let us assume that the Foo Company has two subsidiaries: Foo United States and Foo Europe. The Foo Company could then create a dimension called **Subsidiary** with the two possible values **Foo United States** and **Foo Europe**. The possible values of an explicit dimension are called **members**.

Both dimensions and members are defined using report elements, just like concepts and abstracts before them. To symbolize that a member belongs to a dimension, the member is defined as a child of the dimension in the definition network.

The members of a dimension can also have even more child members themselves. For example, the Subsidiary Foo Europe could have two subsidiaries: Foo Switzerland and Foo EU.

Figure 4.17: Visualizations of the explicit dimension "Subsidiary"



4.8.3 Typed dimensions

Typed dimensions are dimensions that do not have a predefined set of possible values. Instead, the values of a typed dimension are constrained by a data type. For example, a dimension could be constrained to only allow values of the type `xs:integer`.

Similar to explicit dimensions, typed dimensions are defined using report elements. Unlike explicit dimensions, typed dimensions do not have members. They consist solely of the dimension report element, which defines the data type of the dimension.

4.8.4 Line items and hypercubes

With our current understanding of hypercubes, we can only view the whole report as a single, gigantic hypercube. Especially when considering the additional dimensions, most facts will use only a small subset of the possible dimensions. This makes the resulting hypercube high dimensional, with most of the dimensions being unused. Using the large hypercube as a basis for analysis would be very inefficient.

To solve this problem, XBRL introduces the **hypercube** report element. Conceptually, a hypercube is a sub-hypercube of the whole report hypercube. Hypercube report elements are usually defined on a per-role basis as part of a definition network. It picks a subset of the dimensions of the whole report hypercube. This subset is determined in the definition network, where "hypercube-dimension" arcs specify which dimensions are part of the hypercube.

Besides the hypercube report element, XBRL also introduces the **lineItems** report element. LineItems are used to specify which concepts are part of the hypercube. Reports can specify tens of thousands of concepts, but only a few of them are relevant for a particular role. The LineItems report element specifies the relevant concepts by listing them as children in the definition network.

If understanding lineItems proves to be difficult, consider the following: LineItems are to concepts what dimensions are to members.

4.9 XBRL Epilogue

As we have seen in this chapter, XBRL is a complex standard with many moving parts. It is a standard that has been in development for over 20 years, and it shows. In the first half of this chapter, we have seen how an XBRL report is, at its core, just a collection of facts. The second half of this chapter has shown us how these facts can be structured into networks, roles and hypercubes. This chapter was in no way exhaustive, and there are many more aspects of XBRL that we have not covered. However, the aspects that we have covered are the ones that are most relevant to this thesis. The chapter was also, as mentioned in the introduction, heavily based on both the XBRL standard[7] created by Charles Hoffman and Dr. Ghislain Fourny's interpretation of it in *the XBRL Book*[6].

The next chapter will introduce the API of Brel, and how it interprets the XBRL standard. The API chapter precedes the implementation chapter, which will explain how Brel implements the XBRL standard. Even though the API is part of the result of this thesis, it makes more sense to introduce the API before its implementation.

Chapter 5

API

Note: I am still actively working on this and subsequent chapters. Proceed with caution.

TODO: write API chapter and generate more UML diagrams.

Chapter 6

Implementation

6.1 Overview

The creation of brel could be divided into three main parts The design of the API, the implementation of the API and testing of the API. The bulk of both the API design and testing will be discussed in chapter 7. This chapter will focus on the implementation of the API.

The implementation of the API was done in the programming language Python. We chose python because it is a high level language and one of the most popular programming languages in the world. Python is also quite popular outside of the computer science and software engineering communities as well. Its simplicity and readability makes it a good choice for beginners and non-programmers.

Brel parses XBRL reports, which are used by companies to report their financial information. So the target audience of brel are accountants and financial analysts. These people are not necessarily programmers, so using a language that is easy to read and understand is just as creating an API that is easy to read and understand. Chapter 4 already discussed the structure of XBRL reports. In summary, XBRL reports are XML documents that contain the following information:

- A flat list of facts
- A list of report elements, many of which are created by external taxonomies
- A system for identifying different report elements, and other objects in the report. Brel uses QNames.
- Components which form the chapters of the report.
- Networks which link report elements together to model relationships between them.

This chapter will cover all the aforementioned elements of XBRL reports in their own section. It will discuss the difficulties of parsing these elements and how brel solves these difficulties. The first section is about QNames, which are used extensively throughout XBRL reports.

6.2 QNames and namespace normalization

6.2.1 Overview

As mentioned in chapter 4, Brel leverages QNames to identify various elements across the XBRL report. QNames are a concept that is widely used in XML and

XML-based languages, such as XBRL. Thus, for most QNames in Brel, the necessary information can be directly extracted from the corresponding XML elements in both the XBRL taxonomy and the XBRL filing. However, there is an important difference between QNames in XML and QNames in Brel - Namespace bindings. Namespace bindings are the mappings between prefixes and namespace URIs. For example, the prefix **us-gaap** might be bound to the namespace URI <http://fasb.org/us-gaap/2023>. In XML documents, namespace bindings can be defined on a per-element basis. Child elements inherit the namespace bindings of their parent elements, unless they define their own namespace bindings. This allows for the construction of complex namespace hierarchies, where each element can have its own namespace bindings. In Brel, however, namespace bindings are flat and defined on a global level. The process of converting this hierarchical structure of namespace bindings into a flat structure is called **namespace normalization**. Namespace normalization not only flattens the namespace hierarchy, but also resolves collisions in namespace bindings. The motivation for having a flat structure of namespace bindings is that it makes it easier for users to search concepts, types, etc. in an XBRL filing using QNames. Lets say a user wants to search for all facts that are associated with the concept **us-gaap:Assets**. In XML, the prefix **us-gaap** might be bound to a different namespace URI in each element. So there are no guarantees that the prefix **us-gaap** is bound to just one namespace URI in the entire XBRL filing. A good XBRL parser will find all namespace bindings for the prefix **us-gaap** and search all of them for the concept **Assets**. However, this is not guaranteed and depends on the XBRL parser. Furthermore, the filer of the XBRL report might choose to use the prefix **us-gaap1** instead of **us-gaap** in some sections to refer to the same namespace URI. So there might not be a concept called **us-gaap:Assets**, but a concept called **us-gaap1:Assets** instead. The user could simply search both for **us-gaap** and **us-gaap1** to find the concept **Assets**. Obviously, this is extremely cumbersome and a suboptimal fix for the problem. It also insufficiently shields users from the complexity of XML and XML-based languages, such as XBRL, since they need to be aware of the structure of the underlying namespace hierarchy. In Brel, the user can simply search for the QName **us-gaap:Assets** and will find all facts that are associated with this concept. Brel will automatically resolve the prefix **us-gaap** to the corresponding namespace URI. If the report uses the prefix **us-gaap1** instead of **us-gaap**, Brel will still be able to resolve the prefix to the correct namespace URI.

6.2.2 Namespace hierarchy notation

This section exclusively focuses on the implementation of QNames in Brel and namespace bindings in particular. Since XML documents tend to be verbose and contain a lot of information that is not relevant for namespace bindings, I will use a custom notation to represent namespace bindings in this section. I will refer to this notation as the *namespace hierarchy notation* and describe it using an example. The namespace hierarchy notation works as follows:

- Each level of the namespace hierarchy is represented as a single line.
- Depending on the level of the namespace hierarchy, the element name has a different indentation.
- The name of the element is followed by a list of namespace bindings, separated by commas.

- Each namespace binding is represented as a key-value pair, where the key is the prefix of the namespace binding and the value is the namespace URI.
- If an element does not define any namespace bindings, the list of namespace bindings is omitted.
- XML attributes that are not namespace bindings are omitted in this notation.

Take the following XML snippet as an example:

Figure 6.1: Example of an XML snippet where namespace bindings are defined on a per-element basis

```
<element1
  xmlns:foo = "http://foo.com"
  color = "red"
>
  <element2 xmlns:bar = "http://bar.com"/>
    <element3 color = "blue">
      </element3>
    </element2>
  </element1>
<element4
  xmlns:baz = "http://baz.com"
  xmlns:foo = "http://other-foo.com"
>
```

Using the namespace hierarchy notation, we can represent the same namespace hierarchy as follows:

Figure 6.2: Example of the same XML snippet in our custom notation

```
root
├─ element1 foo = "http://foo.com"
│   └─ element2 bar = "http://bar.com"
│       └─ element3
│           └─ element4 baz = "http://baz.com", foo = "http://other-foo.com"
```

As we can see, the namespace hierarchy notation is much more compact than the XML snippet. The `color` attribute of `element1` and `element3` is omitted, since it is not a namespace binding.

6.2.3 Flattening namespace bindings

As mentioned in the previous section, namespace bindings in XML are defined on a per-element basis, which is arranged in a tree structure. In Brel, however, namespace bindings are defined on a global level, which is arranged in a flat structure. Thus, we need to flatten the namespace hierarchy of the XBRL taxonomy into a flat structure.

The process of flattening a tree structure into a flat structure is a common problem in computer science. One of the most common approaches to this problem is to use a depth-first search algorithm. This is also the approach that I use in Brel to flatten the namespace hierarchy of the XBRL taxonomy.

Remember that in XML, child elements inherit the namespace bindings of their parent elements. Thus, when flattening the namespace hierarchy, we need to make

sure that all the namespace bindings of a parent are also present in its children, unless the children define their own namespace bindings. To give an example of flattening, let us flatten the namespace hierarchy from the previous figure 6.2

Figure 6.3: Example of the same XML snippet in our custom notation, flattened

```
root
├─ element1 foo = "http://foo.com"
├─ element2 foo = "http://foo.com", bar = "http://bar.com"
├─ element3 foo = "http://foo.com", bar = "http://bar.com"
└─ element4 baz = "http://baz.com", foo = "http://other-foo.com"
```

As we can see, the namespace hierarchy has been flattened into a flat structure ¹, meaning that all elements are on the same level. The order of the elements is determined by the depth-first search algorithm.

Each child element inherits the namespace bindings of its parent element. Thus, the `element2` and `element3` elements inherit the namespace bindings of the `element1` element.

To extract the namespace bindings of this flat structure, we can simply iterate over the elements and extract the namespace bindings of each element. For our example, this would result in the following list of namespace bindings:

Figure 6.4: List of namespace bindings extracted from the flattened namespace hierarchy

```
foo = "http://foo.com"
bar = "http://bar.com"
baz = "http://baz.com"
foo = "http://other-foo.com"
```

6.2.4 Collisions in namespace bindings

An observant reader might have noticed that the list of namespace bindings from the previous section contains two bindings for the `foo` prefix. The first binding is defined as `foo = "http://foo.com"`, whereas the second binding is defined as `foo = "http://other-foo.com"`.

This is called a collision. Brel prohibits most collisions in namespace bindings, but allows some of them. The following section describes the different types of collisions and how they are handled in Brel.

6.2.5 Types of namespace collisions

There are three types of namespace collisions that can occur in Brel:

- **Version collision:** Two namespace bindings have the same prefix and the same namespace URI, but different versions of it. The version is defined as the numbers and dashes in the namespace URI that indicate if the URI is more recent than another URI.

Example: `foo = "http://foo.com/2022"` and `foo = "http://foo.com/2023"`

¹Technically, the namespace hierarchy is not completely flat, since the root element is still present. However, the root element does not contain any namespace bindings, so it does not affect the namespace hierarchy.

- **Prefix collision:** Two namespace bindings have the same prefix, but different *unversioned* namespace URIs. A unversioned namespace URI is a namespace URI with all version information removed.

Example: `foo = "http://foo.com"` and `foo = "http://other-foo.com"`

- **Namespace URI collision:** Two namespace bindings have the same *unversioned* namespace URI, but different prefixes.

Example: `foo = "http://foo.com"` and `bar = "http://foo.com"`

Version collision

Version collisions occur when two namespace bindings have the same prefix and the same namespace URI, but different versions of it.

Version collisions are allowed in brel, but they do raise an interesting question: Lets say the creates a QName `foo:bar` to search for a concept in the taxonomy. Also assume that the XBRL filing contains the following namespace bindings:

Figure 6.5: Example of a version collision

```

root
├── element1 foo = "http://foo.com/01-01-2022"
│   └── foo:bar
└── element2 foo = "http://foo.com/01-01-2023"
    └── foo:baz

```

Which namespace URI should be used for the QName `foo:bar`?

The mechanism that Brel implements is straightforward: Use the newest version. First, Brel will remove all digits, dashes and dots from the URI versions. If the two URI versions are equal after this step, then they are considered the same URI with different versions. In our example, both URI versions transform into `http://foo.com/`.

Second, for each URI version, Brel will extract all numbers and compute their sum. The URI version with the higher sum is considered the newer version. For our example, the sum of the first URI version is 2024, whereas the sum of the second URI version is 2025. Thus, the second URI version is considered the newer version. So if the user searches for the QName `foo:bar`, the URI version `http://foo.com/01-01-2023` will be used.

Even though this mechanism is straightforward, it does have some drawbacks. Namely, theoretically it is easy to trick the mechanism into using an older version of a namespace URI. For example, the URI version `http://foo.com/31-12-2021` would be considered newer than `http://foo.com/01-01-2023`.

However, this is not a problem in practice since version collisions tend to be rare. On top of that, most taxonomies are released on a yearly basis and their URI just contains the year of the release. If the URI only contains the year, then the mechanism works as expected.

Furthermore, the mechanism used for searching and comparing QNames in Brel only uses the prefix and the local name of the QName. Therefore, even if the mechanism would be tricked into using an older version of a namespace URI, it would not affect the search results.

Prefix collision

A prefix collision occurs when two namespace bindings have the same prefix, but different *unversioned* namespace URIs. The following figure shows an example of a

prefix collision:

Figure 6.6: Example of a prefix collision

```
root
├─ element1 foo = "http://foo.com"
│   └─ foo:bar
├─ element2 foo = "http://other-foo.com"
│   └─ foo:baz
```

Prefix collisions are not allowed in Brel. Brel will rename one of the prefixes to avoid the collision. In the case of our example above, Brel will `element2`'s binding to `foo1 = "http://other-foo.com"` and will replace all appropriate QNames with the new prefix.

Figure 6.7: Example of a resolved prefix collision

```
root
├─ element1 foo = "http://foo.com"
│   └─ foo:bar
├─ element2 foo1 = "http://other-foo.com"
│   └─ foo1:baz
```

If the user searches for a QName `foo:bar`, Brel will both search for `foo:bar` and `foo1:bar`.

Namespace URI collision

A namespace URI collision occurs when two namespace bindings have the same *unversioned* namespace URI, but different prefixes. An example of a namespace URI collision is shown in the following figure:

Figure 6.8: Example of a namespace URI collision

```
root
├─ element1 foo = "http://foo.com"
│   └─ foo:bar
├─ element2 bar = "http://foo.com"
│   └─ bar:baz
```

Namespace URI collisions are not allowed in Brel. Brel will pick one of the two prefixes as the preferred prefix and will rename the other prefix to avoid the collision. In general, Brel will pick the shorter prefix as the preferred prefix. If both have the same length, Brel will pick the prefix that comes first alphabetically. In the case of our example, `bar` will be picked as the preferred prefix. Brel will rename the prefix `foo` along with all occurrences to `bar`.

Figure 6.9: Example of a resolved namespace URI collision

```
root
├─ element1 bar = "http://foo.com"
│   └─ bar:bar
├─ element2 bar = "http://foo.com"
│   └─ bar:baz
```

There are some prefixes that are considered special and will always be picked as the preferred prefix, regardless of their length or alphabetical order. These special prefixes do not even have to be defined in the XBRL taxonomy. If there is a namespace binding that points to the same namespace URI as one of the special prefixes, the special prefix will be picked as the preferred prefix. The following prefixes are considered special:

- `xml` = `"http://www.w3.org/XML/<year>/namespace"`
- `xlink` = `"http://www.w3.org/<year>/xlink"`
- `xs` = `"http://www.w3.org/<year>/XMLSchema"`
- `xsi` = `"http://www.w3.org/<year>/XMLSchema-instance"`
- `xbrli` = `"http://www.xbrl.org/<year>/instance"`
- `xbrldt` = `"http://xbrl.org/<year>/xbrldt"`
- `link` = `"http://www.xbrl.org/<year>/linkbase"`
- `xl` = `"http://www.xbrl.org/<year>/XLink"`
- `iso4217` = `"http://www.xbrl.org/<year>/iso4217"`
- `utr` = `"http://www.xbrl.org/<year>/utr"`
- `nonnum` = `"http://www.xbrl.org/dtr/type/non-numeric"`
- `num` = `"http://www.xbrl.org/dtr/type/numeric"`
- `enum` = `"http://www.xbrl.org/year/extensible-enumerations"`
- `enum2` = `"http://www.xbrl.org/PR/<date>/extensible-enumerations-2.0"`
- `formula` = `"http://www.xbrl.org/<year>/formula"`
- `gen` = `"http://www.xbrl.org/<year>/generic"`
- `table` = `"http://www.xbrl.org/<year>/table"`
- `cf` = `"http://www.xbrl.org/<year>/filter/concept"`
- `df` = `"http://www.xbrl.org/<year>/filter/dimension"`
- `ef` = `"http://www.xbrl.org/<year>/filter/entity"`
- `pf` = `"http://www.xbrl.org/<year>/filter/period"`
- `uf` = `"http://www.xbrl.org/<year>/filter/unit"`
- `ix` = `"http://www.xbrl.org/<year>/inlineXBRL"`
- `ixt` = `"http://www.xbrl.org/inlineXBRL/transformation/<date>"`
- `entities` = `"http://xbrl.org/entities"`

If, for example, the XBRL filing contains a namespace binding `foo` = `"http://www.w3.org/2001/XMLSchema-in"` then the prefix `xsi` will be picked as the preferred prefix and all occurrences of `foo` will be renamed to `xsi`.

The special prefixes and the corresponding namespace URIs can be configured in

in the `nsconfig.json` file.

With this knowledge of QNames and namespace normalization, we can now move on to the next section which explains how Facts are implemented in Brel. Facts use QNames for a few of their characteristics, most notably their concept characteristic.

6.3 Implementation of Facts

6.3.1 Overview

Facts in Brel function very similar to facts in XBRL. However, there are a couple of key differences between XBRL and Brel that are worth highlighting. Before we dive into the details of facts in Brel, let us first look at how facts are represented in XBRL.

6.3.2 Facts in XBRL

In the context of XBRL, facts are represented as XML elements in an XBRL instance document. The following snippet shows an example of a fact in XBRL:

Figure 6.10: Example of a fact in XBRL

```
<us-gaap:ExtinguishmentOfDebtAmount
  contextRef="c-216"
  decimals="-6"
  id="f-727"
  unitRef="usd">
  121000000
</us-gaap:ExtinguishmentOfDebtAmount>
```

1

Thinking back to how we defined facts in chapter TODO, we can see that the above example does not contain all the information that we defined as being necessary for a fact. In particular, the above example does only contain information about the concept, the value, and the unit of the fact. Both the concept and the unit are merely links to other elements in the XBRL taxonomy, but do not contain any information about the concept or the unit themselves. Furthermore, instead of providing information about the entity and the period, the example only contains a reference to a context element.

6.3.3 Contexts in XBRL

In XBRL, contexts are used to provide information about the entity and the period of a fact. If the fact is part of a hypercube, the context also contains information about the dimensions and members of the fact. Contexts are defined in the instance document using the `context` element.

Going back to our example from above, the context element referenced by the fact is defined as follows:

Figure 6.11: Example of the context referenced by the fact in figure 6.10

```
<context id="c-216">
  <entity>
    <identifier scheme="http://www.sec.gov/CIK">
      0000021344
    </identifier>
    <segment>
      <xbrldi:explicitMember dimension="dei:LegalEntityAxis">
        ko:BottlingOperationsInAfricaMember
      </xbrldi:explicitMember>
    </segment>
  </entity>
  <period>
    <startDate>2023-01-01</startDate>
    <endDate>2023-06-30</endDate>
  </period>
</context>
```

2

The XML snippet fills in some of the missing information from the fact, namely:

- The **entity** refers to *who* is reporting. The entity XML element provides us with a identifier for the entity, as well as information about where the identifier comes from. In this case, the identifier is a Central Index Key (CIK) provided by the US Securities and Exchange Commission (SEC). If we look up the identifier "0000021344" in the SEC's database, we can see that it corresponds to the Coca Cola Company.
- The **period** refers to *when* the information is being reported. As described in chapter TODO, a period can either be a point in time or a duration. In this case, the period is defined as a duration, starting on January 1, 2023 and ending on June 30, 2023.
- The **dimensions** refer to additional information about the fact. This information is only relevant if the fact is part of a hypercube. Dimensions are defined as child elements of the **segment** element, which is a child element of the **entity** element.

A observant reader might have noticed that the placement of dimensions in the context element is not consistent with the placement of both the entity and the period information of a fact. The main reason for this inconsistency is that the XBRL specification was not designed with hypercubes in mind. Hypercubes were added to the XBRL specification at a later point in time.

A second oddity of facts in XBRL is that they are not self-contained. Their definition is spread across multiple elements in the instance document. Most noticeably, the context element is defined separately from the fact element. Since different XBRL facts reference the same context element, the incentive behind this design decision is clear: Reducing redundancy. A XBRL instance document can contain thousands of facts, but only a handful of context elements.

However, this design falls apart as soon as we consider the case of hypercubes. To illustrate the problem, consider the following numbers:

In their 2023 Q2 report, the Coca Cola Company

1. reported 1658 facts.

2. defined 13 context elements without dimensions.
3. defined 397 context elements with dimensions.
4. used the same entity for all facts.
5. used around 15 different periods for all facts.

So the vast majority of all contexts exist to provide information about the dimensions of a fact. However, the entity and the period information are mostly redundant.

6.3.4 Units in XBRL

In the same way that contexts are used to provide information about the entity and the period of a fact, units are used to provide information about the unit of a fact. Units are also defined in the instance document using the `unit` element. Going back to our example from above, the unit element referenced by the fact is defined as follows:

```
<unit id="usd">
  <measure>iso4217:USD</measure>
</unit>
```

3

This XML snippet connects our intuitive understanding of the unit "usd" with a formal definition of the unit. It does so by referencing the official ISO 4217 currency code for the US dollar.

6.3.5 Concepts in XBRL

As mentioned in chapter 4, concepts are the fundamental building blocks of XBRL. Each fact is associated with exactly one concept. Unlike units, periods, and contexts, concepts are not defined in the instance document. Instead, they are defined in the XBRL taxonomy, which is a collection of XML schema documents. In case of the above example, the concept is defined in the US GAAP taxonomy, which is a collection of XBRL taxonomies for US Generally Accepted Accounting Principles (GAAP). This already gives us a hint about where to look for the definition of the concept.

The following snippet shows the definition of the concept in the US GAAP taxonomy:

```
<xs:element id="us-gaap_ExtinguishmentOfDebtAmount" name="ExtinguishmentOfD
```

4

The main purpose of this snippet is to provide a formal definition of the concept and to constrain the values that are allowed for the concept. In particular, it tells us that the concept is a monetary item, that it has a debit balance, and that it has a duration period type.

6.3.6 Facts in Brel

In Brel, facts are represented as Python objects. When processing facts in XBRL, the information about the fact is spread across multiple elements in multiple documents. The main goal of Brel is to provide a more intuitive interface for working with facts. Whereas XBRL facts in XML, concepts, units, periods, etc. are all treated differently, Brel facts treat all aspects of a fact equally.

³taken from COCA COLA CO's 2023 Q2 report

⁴taken from the US GAAP taxonomy

6.4 Components Implementation

6.4.1 Overview

Chapter 4 introduced the concept of components in XBRL. To recap, components act as chapters of a report and link together facts that belong to the same chapter. In this chapter, we will look at how components are implemented in Brel.[8] Brel closely follows the XBRL definition of components.¹ Throughout this chapter, we will use a running example to illustrate the concepts introduced in this chapter. The XBRL specification uses the term "role" while Brel uses the term "component" to refer to the same concept.

6.4.2 Components in XBRL

Components in XBRL are defined using the `link:roleType`² element. The following snippet shows an example of a component in XBRL:

Figure 6.12: Example of the component "Inventory" in XBRL from Tesla Inc.'s 2023 Q3 report[8]

```
<link:roleType id="Inventory" roleURI="http://www.tesla.com/role/Inventory">
  <link:definition>0000010 - Disclosure - Inventory</link:definition>
  <link:usedOn>link:presentationLink</link:usedOn>
  <link:usedOn>link:calculationLink</link:usedOn>
  <link:usedOn>link:definitionLink</link:usedOn>
</link:roleType>
```

I will refer to the component in figure 6.12 as the "Inventory" component. All properties of XBRL components are either required or optional. I will use the notation (**required**) and (**optional**) to indicate the requiredness of a property. The Inventory component has the following properties:

- **roleURI** (required): The URI of the Inventory component. This URI is used to reference the component from other elements in the XBRL taxonomy. It is the primary identifier of the component. The roleURI has to be unique on a per-taxonomy basis.[13]
- **id** (optional): The id of the component, which is commonly used to reference the component from other elements in the XBRL report.^{3 4}
- **definition** (optional): A human-readable description of the component.
- **usedOn**: A list of links that the component is used in. Whenever a link uses a component, it must declare the component in the **usedOn** property.⁵ Therefore, the **usedOn** property can be interpreted as a list of possible links that can use the component. However, even though the Inventory component indicates the use of e.g. a presentation link, it does not mean that any presentation ;oml is defined for the component.

¹At the time of writing, the Open Information Model (OIM) does not define components.

²The prefix "link" commonly refers to the linkbase namespace 'http://xbrl.org/2003/linkbase'.

³Even though the id is optional, it is recommended to provide an id for each component.[13]

⁴Similar to the roleURI, the id has to be unique on a per-taxonomy basis.[30]

⁵The **usedOn** property is only required for standard extended links and standard resource elements.[13]

Most of the roles in XBRL will not be defined by the creators of the report, but rather by the creators of the XBRL taxonomy. In the case of Tesla Inc.'s 2023 Q3 report, only 60 components are defined by the creators of the report, while roughly 600 components are defined by the creators of the XBRL taxonomy.[8]

6.4.3 Components in Brel

The main difference between components in XBRL and Brel is that Brel abstracts away the `usedOn` property. Instead, components in Brel provide a direct way of accessing the links that use the component. A component in Brel is defined using the `Component` class. The following UML diagram shows the class diagram of the `Component` class:

Figure 6.13: UML diagram of the `Component` class

brel.brel_component.Component
<code>get_URI(): str</code> <code>get_calculation_network(): CalculationNetwork None</code> <code>get_definition_network(): DefinitionNetwork None</code> <code>get_info(): str</code> <code>get_presentation_network(): PresentationNetwork None</code> <code>has_calculation_network(): bool</code> <code>has_definition_network(): bool</code> <code>has_presentation_network(): bool</code> <code>is_aggregation_consistent(facts: list[Fact]): bool</code>

The `get_info` method returns the text of the definition element if it exists, otherwise it returns `None`. This renaming was done to avoid confusion with the `get_definition_network` method that returns the definition network of the component.

The `get_URI` method returns the URI of the role of the component.

Even though the `usedOn` property has been discarded, during parsing, Brel will still perform a check for the `usedOn` property. More precisely, if there exists a non-standard network that is associated with the component, Brel will check if the `usedOn` property is defined for the component. If not, Brel will raise an exception. The three methods `get_presentation_network`, `get_calculation_network`, and `get_definition_network` return the respective networks of the component. They return `None` if the network in question does not exist for the component.

Using the three methods `has_presentation_network`, `has_calculation_network`, and `has_definition_network`, it is possible to check if the respective network exists for the component.

Finally, the `sums_up` method acts as a shortcut for the `sums_up` of the calculation network of the component. If the component does not have a calculation network, the `sums_up` method will return `True`.

6.5 Implementation of Networks

6.5.1 Overview

In chapter 4, we introduced the concept of networks. We also looked at the different types of networks that are defined in the XBRL 2.1 specification. In this chapter, we will look at how networks are represented in XBRL and how Brel parses them.

As previously mentioned, one of the main goals of Brel is to shield the user from the complexity of the XBRL specification. Brel achieves this on two different levels.

First, Brel provides a simple API that allows users to access and traverse networks independent of the type of network. This API provides methods for traversing any directed acyclic graph. It is intended for advanced users and is useful for debugging networks in a report.

For example, for each network, Brel provides a function that returns all direct children of a node in the network.

Second, Brel implements a wrapper around each type of network. This wrapper provides a clean interface for some of the most common operations on networks. The functionality of the wrapper depends on the type of network.

For example, the wrapper for calculation networks provides functions to validate the calculation network and to calculate the value of a concept.

We will first look at the common interface for all networks. Then, we will look at the different types of networks and see how they diverge from the common interface.

6.5.2 Common Interface

A network in Brel comprises of two different classes: **INetwork** and **INetworkNode**. Python does not support interfaces, so we used abstract base classes to define the interface of networks in Brel.

A network in brel is a directed acyclic graph.² Each node in the graph has an ordered list of children and at most one parent. The network does not necessarily have to be connected, i.e. there can be multiple disconnected subgraphs in the network.

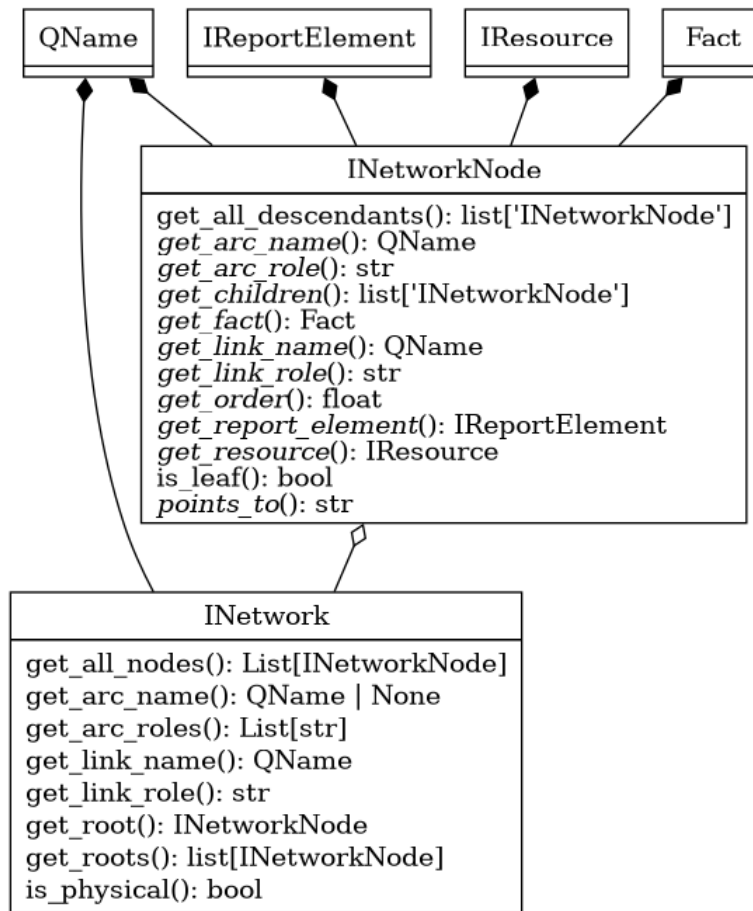
Technically speaking, a network in Brel is just a set of root nodes. These nodes point to their children, which in turn point to their children, and so on. So, to traverse a network, we only need to know the root nodes of the network. The **INetwork** class provides a method to get all root nodes of the network and the **INetworkNode** class provides a method to get all children of a node, which allows us to traverse the network.

One implementation detail of networks in Brel is that they cannot be empty. They have to contain at least one node. The functionality of an "empty" network can still be achieved by letting the component of the network point to **None**.

The following UML diagram shows the class diagram of the **INetwork** class. The classes **QName**, **IReportElement**, **IResource**, and **Fact** are depicted in a simplified form.

²The XBRL specification does allow cycles in networks, but Brel does not support them.

Figure 6.14: Class diagram of the `INetwork` and `INetworkNode` classes



First, we will look at the methods of the `INetwork` class.

INetwork

The `INetwork` class provides the following methods:

- **get_roots:** Returns a list of all root nodes of the network.
- **get_root:** Returns the root node of the network if the network has exactly one root node, otherwise it raises an exception.
- **get_all_nodes:** Returns a list of all nodes that are part of the network i.e. all nodes that are reachable from the root nodes.
- **get_link_role:** Returns the link role of the network. The link role is usually the URI of the component that the network belongs to. If the network does not belong to a component, it returns the standard link role <http://www.xbrl.org/2003/role/link>.
- **get_link_name:** Returns the link QName of the network. This can be `link:presentationLink`, `link:calculationLink`, etc.
- **get_arc_roles:** Returns all the different arc roles that are used by nodes in the network. For most networks, this will only be one arc role. For example, for

a calculation network, this will be `summation-item`. However, for a definition network, multiple arc roles can be used within the same network.

- **get_arc_name:** Returns the arc QName of all arcs in the network. All arcs in the network must have the same arc name. For example, for a calculation network, this will be `link:calculationArc`. Since a network can consist of a single node, it is possible that the network does not contain any arcs. In that case, the method returns `None`.
- **is_physical:** Returns `True` if the network is a physical network, otherwise it returns `False`. For all physical networks, the whole network can only have a single arc role.

INetworkNode

The `INetworkNode` class provides the following methods:

- **get_children:** Returns a list of all direct children of the node. The children are sorted by their order attribute.
- **get_get_all_descendants:** Returns a list of the transitive closure of the children of the node.
- **get_order:** Returns the order attribute of the node.
- **is_leaf:** Returns `True` if the node has no children, otherwise it returns `False`.
- **get_link_role:** Returns the link role of the network that the node belongs to.
- **get_link_name:** Returns the link QName of the network that the node belongs to.
- **get_arc_role:** Returns the arc role of the arc that connects the node to its parent.
- **get_arc_name:** Returns the arc QName of the arc that connects the node to its parent.
- **points_to:** Returns the type of element that the node points to. This can either be 'report element', 'resource' or 'fact'.
- **get_report_element:** Returns the report element that the node points to. Raises an exception if the node does not point to a report element.
- **get_resource:** Returns the resource that the node points to. Raises an exception if the node does not point to a resource.
- **get_fact:** Returns the fact that the node points to. Raises an exception if the node does not point to a fact.

Given these two interfaces, we can now discuss how Brel parses links into networks.

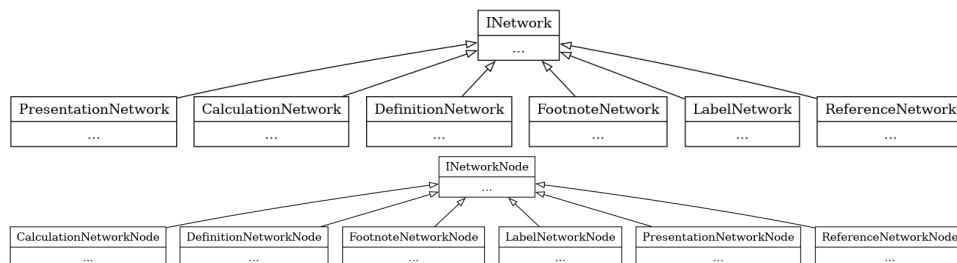
6.5.3 Parsing Links into Networks

Section 4.6 introduced the different types of networks, and how they are all just a collection of arcs, locators and resources. It also explained how locators and resources represent nodes and how arcs represent edges in a network. So parsing a link into a network is just a matter of converting a list of nodes and edges into a graph. Brel does this in four steps:

1. First, it iterates over all elements of the link. If the element is a locator or a resource, it creates an instance of the **INetworkNode** class. If the element is an arc, Brel stores the from and to attributes of the arc in an edge list.
2. For the second step, Brel can assume that all nodes have been created. So, it iterates over the edge list and adds the to-node as a child to the from-node.
3. Next, Brel iterates over all nodes and adds all nodes that do not have a parent to the root list of the network. The root list is then wrapped into a **INetwork** instance.
4. Finally, if the network has any implications on the report as a whole, brel will apply these implications to the report. For example, if the label network adds a label to a concept, Brel will add the label to the concept in the report. The exact implications depend on the type of network.

So far, we have only looked at the common interface of networks. In reality, there are different types of networks that have different functionality. For each type of network, Brel provides both a node class and a network class, all of which inherit from the **INetwork** and **INetworkNode** classes respectively.

Figure 6.15: Inheritance diagram of the different network and node classes



There exists a factory method for each type of network or node that returns the correct network or node class for the given link. These factory methods are used by the algorithm 6.5.3 to create the correct network and node instances.

6.5.4 Resolving Locators

As mentioned in section 4.6, locators are used to point to report elements or facts. This section will explain how Brel resolves locators.

XBRL locators use so called "XPointer" expressions to reference other XML elements, which are potentially defined in other XML documents.[29] XPointers use the "id" attribute of XML elements to reference them. So, to resolve a locator, Brel first needs to find the XML element that the locator points to. Brel does this by remembering the id of each fact and report element that it has parsed. It builds a dictionary that maps ids to facts and report elements. Then, when it encounters a locator, it uses the id of the locator to look up the fact or report element that the locator points to.

6.5.5 Parsing Resources

Resources are the other type of element that can be referenced by arcs. Unlike locators, resources are not used to point to other elements in the report. Instead, they directly contain the value of the element that they represent.

The current specification of XBRL 2.1 defines three built-in types of resources: label, reference and footnote. Custom resources are also possible.

Resources consist of three elements: a role, a label and a value.

The role is a URI that identifies the type of resource. For example, the role `http://www.xbrl.org/2003/role/terseLabel` identifies a label resource that contains a short label for a concept, whereas the role `http://www.xbrl.org/2003/role/footnote` identifies a footnote resource that contains a footnote for a concept.

The label acts as an identifier for the resource. The label is used by arcs to reference the resource. Do not confuse the label of a resource with a human readable label for a concept.

The value of a resource contains the actual resource. For labels, the value is the text of the label. For references, this is the dictionary that points to some external resource like an article in the SEC's Code of Federal Regulations.

Since resources have all the information that is needed to parse them, Brel does not need to resolve any references to parse them. Therefore, parsing them is very straightforward.

6.5.6 Implications of Networks

As mentioned in section 6.5.3, networks can have implications on the report as a whole. There are two implications that are common to all networks: Labels and LineItems promoting.

6.5.7 Add Labels to Report Elements

Labels are used to provide human-readable names for report elements. They are created using the label network `link:labelLink`. Labellinks are described in more detail in section 4.6.6. Brel parses report elements before it parses networks, since many networks contain locators that point to report elements. So, when Brel parses a label network, it already knows all report elements that are referenced by the locators in the network.

Once the label network has been parsed, Brel iterates over all labels in the network. It then adds the label to the report element that the locator of the label points to.

6.5.8 LineItems Promoting

Report elements are defined in the taxonomy and there six different types: concepts, abstracts, lineitems, members, dimensions and hypercubes. When looking at an XML element in the taxonomy, it is not always clear what type of report element it represents. Both abstracts and lineitems are represented by a structurally identical XML element. There are two approaches to tell them apart:

1. The first approach is to look at the QName of the element. Lineitems often have the substring "LineItems" in their QName. The problem with this approach is that this is more of a convention than a rule. There is no guarantee that the QName of a lineitem will contain the substring "LineItems".
2. The second approach is to look at how the element is used in networks. specifically, definition networks define the relationships between report elements. The arc role `hypercube-dimension` is used to define the relationship between

a hypercube report element and a dimension report element. Similarly, the arc role **all** is used to define the relationship between a hypercube report element and a lineitem report element.

Brel uses the second approach to tell lineitems and abstracts apart. Initially, Brel assumes that all report elements are abstracts. Then, when it parses a definition network, it looks at the arc roles of the arcs in the network. If the network contains an arc with the arc role **all** and the arc lies between a hypercube and an abstract, Brel assumes that the abstract is a lineitem.

Chapter 7

Results

TODO: Write this chapter

Chapter 8

Conclusion

TODO: Write this chapter

Bibliography

- [1] European Banking Authority. Eba xbrl filing rules. <https://extranet.eba.europa.eu/sites/default/documents/files/documents/10180/2185906/63580b57-b195-4187-b041-5d0f3af4e342/EBA%20Filing%20Rules%20v4.3.pdf?retry=1>, 2018.
- [2] European Banking Authority. Reporting frameworks. <https://www.eba.europa.eu/risk-and-data-analysis/reporting-frameworks>, unknown.
- [3] Richard Smith Bruno Tesnière and Mike Willis. *the journal*. price waterhouse coopers, 2002.
- [4] Coca-Cola Company. 10-q (quarterly report) for quarter ending june 27, 2023. <https://www.sec.gov/Archives/edgar/data/21344/000002134423000048/0000021344-23-000048-index.html>, 2023.
- [5] Microsoft Corporation. Annual report 2022. <https://www.microsoft.com/investor/reports/ar22/index.html>, 2022.
- [6] Dr. Ghislain Fourny. *The XBRL Book: Simple, Precise, Technical*. Independently published, 2023.
- [7] Charles Hoffman. Extensible business reporting language (xbrl). <https://www.xbrl.org/>, 2000.
- [8] Tesla Inc. 10-q quarterly report for quarter ending september 30, 2023. <https://www.sec.gov/ix?doc=/Archives/edgar/data/1318605/000162828023034847/tsla-20230930.htm>, 2023.
- [9] XBRL International Inc. Extensible business reporting language (xbrl) 2.1. <https://www.xbrl.org/Specification/XBRL-2.1/REC-2003-12-31/XBRL-2.1-REC-2003-12-31+corrected-errata-2013-02-20.html>, 2003.
- [10] XBRL International Inc. Extensible business reporting language (xbrl) dimensions 1.0. <https://www.xbrl.org/specification/dimensions/rec-2012-01-25/dimensions-rec-2006-09-18+corrected-errata-2012-01-25-clean.html>, 2006.
- [11] XBRL International Inc. Extensible business reporting language (xbrl) generic links 1.0. <https://www.xbrl.org/specification/gnl/rec-2009-06-22/gnl-rec-2009-06-22.html>, 2009.
- [12] XBRL International Inc. Extensible business reporting language (xbrl) 2.1 - 5.1.1 concept definitions. http://www.xbrl.org/specification/xbrl-2.1/rec-2003-12-31/xbrl-2.1-rec-2003-12-31+corrected-errata-2013-02-20.html#_5.1.1, 2013.

- [13] XBRL International Inc. Extensible business reporting language (xbrl) 2.1 - 5.1.3 defining custom role types - the roletype element. http://www.xbrl.org/specification/xbrl-2.1/rec-2003-12-31/xbrl-2.1-rec-2003-12-31+corrected-errata-2013-02-20.html#_5.1.3, 2013.
- [14] XBRL International Inc. Extensible business reporting language (xbrl) 2.1 - 5.2.2 the labellink element. http://www.xbrl.org/specification/xbrl-2.1/rec-2003-12-31/xbrl-2.1-rec-2003-12-31+corrected-errata-2013-02-20.html#_5.2.2, 2013.
- [15] XBRL International Inc. Extensible business reporting language (xbrl) 2.1 terminology - 1.4 terminology (non-normative except where otherwise noted). http://www.xbrl.org/specification/xbrl-2.1/rec-2003-12-31/xbrl-2.1-rec-2003-12-31+corrected-errata-2013-02-20.html#_1.4, 2013.
- [16] XBRL International Inc. About xbrl us. <https://xbrl.us/home/about/>, 2021.
- [17] XBRL International Inc. Open information model 1.0. <https://www.xbrl.org/Specification/oim/REC-2021-10-13+errata-2023-04-19/oim-REC-2021-10-13+corrected-errata-2023-04-19.html#term-component>, 2021.
- [18] AICPA Karen Kernan. Xbrl - the story of our new language. <https://us.aicpa.org/content/dam/aicpa/interestareas/frc/accountingfinancialreporting/xbrl/downloadabledocuments/xbrl-09-web-final.pdf>, 2009.
- [19] Mark V Systems Limited. Arelle - open source xbrl platform. <https://arelle.org/arelle/>, 2010.
- [20] Robin Schmidiger. Brel - business reporting extensible library. <https://github.com/BrelLibrary/brel>, 2021.
- [21] European Securities and Markets Authority (ESMA). filings.xbrl.org. <https://filings.xbrl.org/>.
- [22] European Securities and Markets Authority (ESMA). Esma publishes esef conformance suite. <https://www.esma.europa.eu/press-news/esma-news/esma-publishes-esef-conformance-suite>, 2020.
- [23] U.S. Securities and Exchange Commission (SEC). Interactive data test suite. <https://www.sec.gov/structureddata/osdinteractivedatatestsuite>.
- [24] U.S. Securities and Exchange Commission (SEC). Edgar. <https://www.sec.gov/edgar/searchedgar/accessing-edgar-data.htm>, 1996.
- [25] U.S. Securities and Exchange Commission (SEC). 17 cfr § 229.402 - (item 402) executive compensation. <https://www.govinfo.gov/app/details/CFR-2011-title17-vol2/CFR-2011-title17-vol2-sec229-402>, 2011.
- [26] U.S. Securities and Exchange Commission (SEC). Inline xbrl. <https://www.sec.gov/structureddata/osd-inline-xbrl.html>, 2018.
- [27] U.S. Securities and Exchange Commission (SEC). 17 cfr § 229.402 - (item 402) executive compensation. [https://www.ecfr.gov/current/title-17/part-229#p-229.402\(v\)\(2\)\(iv\)](https://www.ecfr.gov/current/title-17/part-229#p-229.402(v)(2)(iv)), 2023.
- [28] XBRL US. Xbrl us xule. <https://xbrl.us/xule/>, 2021.

- [29] World Wide Web Consortium (W3C). Xml, xlink and xpoiner. https://www.w3schools.com/xml/xml_xlink.asp, 1999.
- [30] World Wide Web Consortium (W3C). xml:id version 1.0. <https://www.w3.org/TR/xml-id/>, 2005.