A Graph Grammar Model of Financial Statements with Heterogeneous Parts

Takeo Yaku\textsuperscript{1}, Koichi Anada\textsuperscript{2}, Koushi Anzai\textsuperscript{3}, Shinji Koka\textsuperscript{1}, Miyuki Shimizu\textsuperscript{1}, and Yuki Shindo\textsuperscript{1}

\textsuperscript{1}Department Computer Science and System Analysis, Nihon University, Tokyo, Japan
\textsuperscript{2}Waseda Research Institute for Science and Engineering, Waseda University, Tokyo, Japan
\textsuperscript{3}Department Economics, Kanto Gakuen University, Gunma, Japan

Abstract - Spreadsheets are frequently used in information processing. In those information processing, automatic generation and automatic verification of spreadsheets are required. However, it is difficult to verify the structure of spreadsheets with flexible calculation ranges and with heterogeneous structures. Accordingly, the modeling of spreadsheets is important. We deal with the modeling of spreadsheets such as financial statements. In this paper, we formalize the structure of financial statements with heterogeneous parts using a context sensitive graph grammar. We propose 62 rewriting rules that provide the spatial order of items of heterogeneous financial statements. Furthermore, we also show derivation of financial statements with the grammar.

Keywords: modeling of spreadsheets, context sensitive graph grammar, financial statements, syntax-directed recognition of logical structures

1 Introduction

The formalization of business documents as in Figure 1 has become an important subject with the progress of e-commerce and e-government (see, e.g. [3]). In order to formalize financial statements, we have to specify the spatial order of items, and specify calculation methods of categorized items. A context sensitive graph grammar was proposed in [5] that specify financial statements only with homogeneous part.

In this paper, we formalize the structure of financial statements with heterogeneous parts using a context sensitive graph grammar (CSGG) (see, e.g. [1]). We first propose rewriting rules that provide the spatial order of items in the financial statements. Next, we also show a derivation process that provides syntax directed recognition process of logical structures.

2 Graph representation of financial statements

We represent financial statements by octgrids [4]. Figures 2 and 3 show a financial statement and its corresponding octgrid. Each node in Figure 3 corresponds to each cell of the financial statement in Figure 2.

![Figure 2. An instance of financial statements in Figure 1.](image1)

![Figure 3. The octgrid for Figure 2.](image2)

3 Graph grammar for heterogeneous parts

Financial statements have the two dimensional grid structures. Accordingly they could not be formalized by context free graph grammars (see, e.g. [2]). Thus, we
construct a CSGG that formalize financial statements with heterogeneous parts (cf. [4]).

3.1 Rewriting Rules

The CSGG for financial statements is a system GGF = (NF, TF, MF, PF, SF), where NF = {S, →, ↓, tr}, TF = {DO, OO, St, DO', OO', St'', ye, ch, ch'', Em, Se, To, To'', pe}, MF = {nwe, swe, ewe, wwe} and SF = S.

The terminal label DO stands for “Domestic Operations”, OO for “Overseas Operations”, St for “Sub total”, ch for “change”, Em for an empty cell, Se for “Section”, To for “Total”, and pe for “perimeter”. Fig. 4 shows a part of the production rules of GGF. Each production rule means the rewriting of the left-hand side graph by the right-hand side graph, where vertices are accompanied by their vertex number.

3.2 Derivation

Following Fig. 5 shows a derivation in GGF. The derivation is consist of following Phases 1–20.

Phase 1: Generate the frame and determine the number of columns.
Phase 2: Change labels of nodes in the 1st and the last line.
Phase 3–4: Generate the 2nd line and change labels of nodes in the 2nd line.
Phase 5–6: Generate the 3rd line and change labels of nodes in the 2nd line into terminal ones.
Phase 7–8: Generate the 4th line and change labels of nodes in the 3rd line into terminal ones.

Phase 9–19: Repeat to generate lines and change labels.
Phase 20: Change labels of nodes in the last line into “perimeter”

![Figure 4. Production rules in GGF.](image)

![Figure 5. A derivation process for Figure 2 by GGF.](image)

4 Conclusion

In this paper, we constructed a CSGG, GGF, which generates graphs for heterogeneous financial statements. GGF has 62 rewriting rules. Furthermore, we showed process. However, it is not verified whether all of heterogeneous financial statements are generated by those 62 rules or not. As future works, we discuss attribute rules that formalize scopes of spread sheet calculation defined by GGF.

We would like to thank Professors G. Akagi and K. Tsuchida for their valuable suggestions.

5 References


