

# On the Reading of Tables of Contents

Prateek Sarkar

Eric Saund

*Palo Alto Research Center, California*

psarkar@parc.com

saund@parc.com

## Abstract

*This paper presents a framework for understanding tables of contents (TOC) of books, journals, and magazines. We propose a universal logical structure representation in terms of a hierarchy of entries, each of which may contain a descriptor and a locator. We enumerate graphical and perceptual cues that provide cues to parsing of tables of contents in terms of this formalism. We make initial suggestions about the form of evaluation metrics for comparing groundtruthed tables of contents with the output of recognition algorithms. Typical and atypical tables of contents are used throughout to illustrate significant phenomena that must be dealt with in principled ways in any general TOC interpretation scheme. Finally we discuss implications of our observations on the design of recognition algorithms.*

## 1 Introduction

Tables of contents (TOC) are guides and informative summaries of books, magazines, journals, reports. Automatic parsing of tables of contents is of great use to many document recognition and information mining applications, but the problem is difficult because of the great variety of formats and styles used. Indeed graphic designers distinguish their work in part by novelty; the space of TOC designs is probably open-ended.

For any particular TOC design, rules can often be engineered to parse TOCs in that or similar designs. Previous work in TOC parsing often relies on strong assumptions about layout constraints of particular publications [12, 9]. But the field seeks a more ambitious goal – to successfully read virtually any book’s TOC, thereby covering a significant part of the design space, including those never seen by the algorithm designer.

Before attempting to devise algorithms to parse tables of contents in general, we must understand the patterns, factors, and constraints that underlie their structure and designs. These will in turn enable the finer specification of the parsing problem, and proposal of metrics for evaluating the success or failure of proposed algorithms. We offer

some initial steps along these lines based on our study of a few hundred tables of contents, a few of which employ as motivating examples.

A TOC is often thought of as a navigational aid, especially in the context of document image analysis. For example, parsing TOCs of published material provides useful cross-linking of content headings and content location during paper to digital conversion. Some approaches exploit this constraint by devising TOC parsing algorithms based on matching items in the TOC with chapter and section headings in the body of the book or article [5, 6, 8].

A table of contents does more than simply aid navigation of a book. It represents a summary organization of thought, content, and rhetorical structure that is rich in information, and can be valuable for semantic indexing, organization, and retrieval of content. Just a study of TOCs of books in a topical domain can provide an intuitive “lay of the land” to a person who is new to the domain. This can be used by digital libraries and online book repositories to organize and present their content. Or, to a person already familiar with a topical domain, a TOC provides insights into the depth of coverage, and variation of emphasis in covered material. People often compare and judge books and volumes by their tables of contents.

To understand and parse tables of contents poses rich intellectual challenges. “Tables” of contents actually turns out to be misnomer. Because they often contain different types of functional components (such as heading, number, author or page-number) tabular structure *can* be a useful means for presentation. But TOCs are actually hierarchical lists describing the rhetorical structure of the book, and true tables are not necessarily always a convenient way to express the multi-level hierarchical structure. As a consequence, tables of contents are seldom simple tables, and often are not tabular at all.

TOCs span a wide range of structure and style. Parsing of TOCs can be challenging and presents problems not yet explored in a principled way. But unlike other structured content presented in visual form (such as tables and diagrams), they can be largely covered by a simple logical representation scheme that we propose. We show, by way of example, that it covers a wide variety of TOCs. On the

basis of this scheme, we also propose metrics to evaluate the performance of automatic TOC parsing systems.

We present no algorithms or proofs, but support the above claims through observation and in-depth consideration of both typical and unusual examples. We consider TOC parsing to be a valuable domain for development of visual structure parsing and functional role labeling.

## 2 A logical representation scheme

The term “table” turns out not to be a particularly useful description of either the logical or the layout structure of many TOCs. The logical structure of a true table suggests that rows and columns generate cells whose meaning is partly derived from their row-column coordinates. One may think that a TOC establishes row-wise correspondence between a descriptor (say chapter heading) and a locator (say a page number). However, not all elements in TOCs necessarily have a page-number or locator.

The layout structure of a true table is designed to make the rows and columns apparent by visual appearance indicators such as alignment, background wash, and graphical separators. However, as we shall see in many of our examples, TOCs do not necessarily have a visual tabular structure.

Tables of contents are indices into the larger documents they precede, many of which are organized into a hierarchical rhetorical structure. Books are often organized into chapters, which are formed of sections, which are formed of subsections. TOCs usually respect this rhetorical structure, as well as the ordering of content. Any summary of content that does otherwise is often given a different name. For example an index summarizes and organizes content by alphabetical, chronological, authoring, or other criteria.

Therefore, we propose that the underlying logical structure of a table of contents is not that of a table at all, but of a hierarchical listing. Formally the structure is a tree, where nodes describe chapters, sections, or rhetorical chunks at other levels of granularity. In the TOC-tree, we shall call each node an *entry*. Each *entry* is a triple: (*descriptor*, *locator*, *children*).

Several investigators have previously identified the hierarchical tree as the appropriate logical model for tables of contents [4, 6, 5], especially in regard to the goal of mapping to XML output. Here, instead of propelling directly to one or another parsing algorithm, we expand on the functional roles comprising such a tree, and their mappings to layout and attribute cues. It is useful to consider diverse examples in order to appreciate the range of ways this logical structure can be expressed, and therefore which any general TOC recognition algorithm must cope with.

A *descriptor* tells about the node. It is often a title or a heading, but can itself be a composite object, such as (*title*, *author*), or (*level-name*, *number*, *heading*), where

*level-name* can be “Chapter”, or “Section” or any name associated with a particular hierarchy level. A *descriptor* may also include a passage of descriptive text, a date, or even pictures.

*Locators* are usually page-numbers, or ranges, but may also include explicit hyperlinks, and volume-labels or other hierarchy information (such as page A.21). *Locators* are useful not only for navigation, but they also provide information about the length to which topics are covered, and thus relative emphasis on different *entry* topics.<sup>1</sup>

*Children* of an *entry* is an ordered list of *entries* that are immediately below this *entry* in the hierarchy. The ordering again is normally in rhetorical order. For example, a chapter entry may have a list of sections as its children. *Children* and *locator* are optional constituents of an *entry*.

The above meta-scheme does not specify the exact schema for *descriptor* or *locator*. Most of the ideas presented here can be developed in this abstraction. Figure 1 illustrates part of a table of contents, and its ground-truth interpretation in our proposed representation. Here we assume that the ground-truth for *descriptor* and *locator* are strings of characters. In an extended representation, decompositions of the descriptor element could be expressed by explicit encoding of such constituents as entry type (e.g. “chapter”, “section”), entry number, entry title, entry author, tagline, etc. Belaid [2] takes this idea to an extreme with part-of-speech tagging approach to TOC recognition.

The identification of document elements corresponding to functional objects can take different forms. For example, to analyze scanned images of tables of contents, the ground-truth encoding for *descriptor* and *locator* could be identifiers for respective regions of support in the image. Such a ground-truth may be useful to test a functional role assignment algorithm that takes a segmented image as input.

We present a few illustrative examples of tables of contents from our collection. In Figure 2(a) the table of contents is a flat list of *entries* with no *children*. The salient columnar structure is used to call out the different functional components of each *entry*. Each *descriptor* is itself subdivided into *author* and *chapter title* which happen to be organized as columns. In Figure 2(b) tabular layout is also used, but at the Chapter level of a hierarchical organization. The “Part” entry omits the locator node, and indentation and font variation are employed to distinguish the functional parts, *title*, and *author*. The complex TOC in Figure 2(c) also uses tabular layout, but alternate chapters are laid out in two mirror image tables that are intertwined. Each *descriptor* consists of a chapter number, an image, a chapter title, and a tagline. The *locator* follows immediately below the *descriptor*, and is in a different font. Figure 7 contains more examples of TOCs readily encoded in our logical scheme, but displaying a variety of stylistic choices.

---

<sup>1</sup>Most electronic TOCs have been robbed of this information because they only offer navigation — hyperlinks in lieu of *locators*.

Chapter 4		
BELIEF UPDATING BY NETWORK PROPAGATION		143
4.1	AUTONOMOUS PROPAGATION AS A COMPUTATIONAL PARADIGM	144
4.1.1	Constraint Propagation and Rule-based Computation*	145
4.1.2	Why Probabilistic Reasoning Seems to Resist Propagation*	148
4.2	BELIEF PROPAGATION IN CAUSAL TREES	150
4.2.1	Notation**	150
4.2.2	Propagation in Chains**	153
4.2.3	Propagation in Trees**	162

```

...
<tocentry>
  <descriptor>
    <str>Chapter 4 BELIEF UPDATING BY NETWORK PROPAGATION</str>
  </descriptor>
  <locator>
    <str>143</str>
  </locator>
  <children>
    <tocentry>
      <descriptor>
        <str>4.1 AUTONOMOUS PROPAGATION AS A COMPUTATIONAL PARADIGM</str>
      </descriptor>
      <locator>
        <str>144</str>
      </locator>
      <children>
        <tocentry>
          <descriptor>
            <str>4.1.1 Constraint Propagation and Rule-based Computation*</str>
          </descriptor>
          <locator>
            <str>145</str>
          </locator>
          <children/>
        </tocentry>
      </children>
    </tocentry>
  </children>
</tocentry>
<tocentry ... (continuation)

</children>
</tocentry>
<tocentry ... (continuation)

</children>
</tocentry>
...

```

Figure 1. Excerpt from a typical table of contents from a technical book (top), and a part of a possible ground-truth assignment.

We call a table of contents *conforming* when it can be meaningfully and correctly interpreted in our logical meta-scheme. Based on our evaluation of several hundred TOCs, we feel comfortable in asserting that documents that are intended to serve as tables of contents but do not conform to our logical scheme are extremely rare.

We also suggest that reasonable people will agree on a unique interpretation of most TOCs in this scheme, and this can be done without having to analyze the book that it summarizes. In other words, the TOC contains enough information for unique human interpretation in most cases. Exceptions are unusual and occur for interesting reasons. One example is Figure 4. Here we observe clear Region entries. It is debatable whether States constitute a proper level of their own, or should be encoded as part of region entries' descriptors. Road entries take their own format pattern, but their ordering is ambiguous. Note that our proposed encoding is one-way: a ground-truth encoding of an observed

Patrizia Violi	<i>Preface</i>	7
Geoffrey Nunberg	<i>Introduction</i>	9
Carla Hesse	<i>Books in time</i>	21
James J. O'Donnell	<i>The pragmatics of the new: Trithemius, McLuhan, Cassiodorus</i>	37
Paul Duguid	<i>Material matters: The past and futurology of the book</i>	63
Geoffrey Nunberg	<i>Farewell to the information age</i>	103

(a)

**PART III: NEUROBIOLOGY**

Neuronal Maps for Sensory-Motor Control in the Barn Owl C. D. Spence, J. C. Pearson, J. J. Gelfand, R. M. Peterson and W. E. Sullivan	366
Models of Ocular Dominance Column Formation: Analytical and Computational Results Kenneth D. Miller, Joseph B. Keller and Michael P. Stryker	375
Modeling Small Oscillating Biological Networks in Analog VLSI Sylvie Ryeckebusch, James M. Bower, and Carver Mead	384

(b)

(c)

Figure 2. (a) A table of contents with a simple tabular layout, and no hierarchy. (b) A TOC from a typical technical publication with nested structure. (c) A less mundane looking table of contents. Functional roles are characterized by font, associations indicated primarily by proximity. A complex columnar structure exists, but page numbers do not have their own column.

table of contents is not required to contain all of the information necessary to reconstruct the page.

**Interfaces:** The recursive tree logical structure readily leads to the definition of *query and manipulation interfaces*. While we do not spell out an interface in detail here, the accessors and functions derive straightforwardly from the tree structure of TOCs' underlying semantics. We list some example interface methods in Figure 5.

This generic interface can lead to models for evaluating automated analysis, and for comparison of documents by tables of contents, even without specifying the exact representation of *descriptor* or *locator*. It is clear, though, that *descriptor* and *locators* should at least allow tests of equality. This interface can also lead to exploration of how to compare books by their TOCs.

We should emphasize here that all this development is

SHORT STORIES – 263	
The Conclusion – 268	
The Raj Seal – 288	
The Lost Jewels – 302	
Purification – 320	
The Parrot's Training – 327	
NOVEL – 331	
FROM <i>The Home and the World</i> – 336	
YOUR ELECTRICAL SYSTEM 4	
Safety First 6	Evaluating Circuits 12
From Power Company to Your House 7	Grounding and Polarization 13
Service Panels 8	Home Safety 14
Household Circuits 10	Inspecting for Safety 15
BASIC SKILLS & MATERIALS 18	
Cable and Wire 20	Stripping Wire 25
Boxes 21	Splicing Wire 26
Switches 22	Attaching Wire to Terminals 27
Receptacles 24	

**Figure 3. The apparent nested tabular structures, actually present hierarchical lists of descriptor-locator pairs.**

possible in the context of the TOC alone, *i.e.*, even without considering correspondence to the document that it summarizes.

### 3 Visual representation

Tables of contents are designed to be read by people, and this fact superimposes two important considerations. First, the human visual system is automatically equipped with certain perceptual abilities that enable it to segment scenes into disparate objects, and to group elements into coherent wholes. The so-called Gestalt rules of perception provide for separating and associating visual objects based on size, color, orientation, alignment, co-circularity, symmetry and common motion, to name the most commonly cited factors. Second, tables of contents are cultural artifacts. Citizens *learn* the conventions that organize of their cultures' writings. Styles evolve, and the reader of 19th century treatises might well be disoriented by the complexity and stylistic liberties taken by a modern magazine layout before they learn what cues to attend to.

The visual representation of tables of contents can be quite varied. Creative designs go beyond a simple tabular layout to convey the organization of content. Perceptual cues and visual structures are used to convey various relationships and differences among different functional parts.

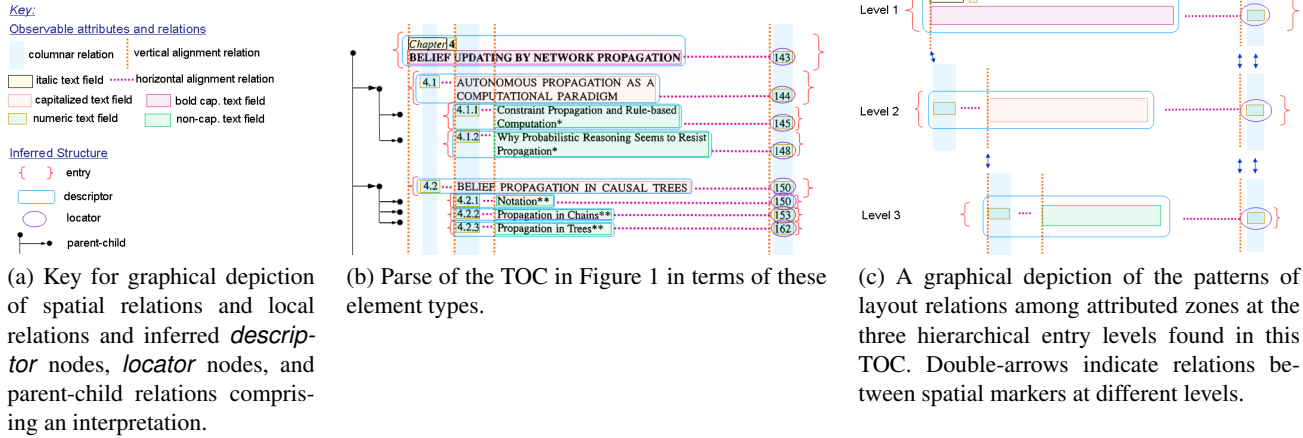
Table 1 lists some kinds of functional roles and relationships commonly found in tables of contents. This listing is not exhaustive of the graphical and perceptual devices that designers employ. Nevertheless, the table illustrates

**Figure 4. A creatively designed TOC which can be associated with ground-truth according to our scheme. However, this ground-truth will fail to explain the rich visual information available to the discerning reader.**

- Entry toc.getRootEntry(), int toc.getDepth(), TOCTree toc.trimDetails(int maxDepthToKeep), TOCTree toc.flatten()
- Descriptor entry.getDescriptor(), Locator entry.getLocator(), TOCTree entry.getSubTOC(), int entry.getLevel(), List entry.getChildren(),
- Entry toc.getMatchingEntry(Descriptor queryDescriptor), List toc.getMatchingEntries(Locator queryLocator)
- Entry entry.getPrecedingEntry(), Entry entry.getPrecedingSibling()

**Figure 5. Examples of interface methods arising from the hierarchical list structure of TOC.**

the maze of ambiguity that the human reader sorts through while navigating a table of contents. It is possible to draw a few dominant themes such as these:



**Figure 6. Spatial relations and local attributes provide evidence for how a table of contents can be parsed into a hierarchy of *descriptor/locator/children* nodes.**

**Table 1. Functional roles and relationships in tables of contents**

Functional implication	Visual/Perceptual cue
Level in hierarchy	Font, size, emphasis, indentation, bullet style, numbering style
Association of descriptor to locator	Alignment, proximity, line-connectors, shared font or color
Functional part delimiter	Column separator, extra white-space, separator lines, color difference, font/size change, bullets, numbering, dashes.
Association of parent to child	Reading order proximity, spatial proximity, relative indentation to child, page-range coverage, emphasis variation through font, size, color, sibling transfer (parent of sibling is parent)
Association between siblings	shared attributes (indentation, font, size, color), reading order proximity, inter-sibling separators.

- If there is visible columnar structure, columns usually contain functionally similar elements such as author or title or page number, while rows comprise our triples. While this is mostly true for left-to-right reading order documents, we present an interesting exception in Figure 8.
- Attribute sharing (alignment, common color, font) supports association, while attribute change indicates different functional chunks.
- A parent to child transition is generally supported by indentation, and/or a reduction in visual saliency (e.g.,

smaller, lighter font).

None of these observations can be translated into definitive rules. No functional phenomenon exclusively entails a perceptual or visual cue. No visual indicator uniquely identifies a functional role or relationship. Rather a logical resolution emerges from simultaneous consideration of a multitude of visual cues.

One source of guidance is the discipline enforced by *style consistency* — any given TOC uses a fixed set of visual cues to indicate any functional role or relationship. For example, a TOC will consistently use dotted-line connectors, or shared font or proximity or any combination of those to indicate association between chapter title and page-number.

Of course, another source of guidance available to humans, that could be made available to algorithms, is familiar patterns or “templates” for tables of contents. The limits of such a strategy are discussed in Section 5.

**Generative exploration:** Starting from a logical tree (hierarchical list) representation, we can imagine composing visual attribute elements and layout style elements to obtain different visual realizations of tables of contents. Some of these will lead to aesthetic and meaningful visuals, others will not. But in principle this would allow the automatic synthesis of tables of contents in the manner of other forms of automatic document layout [1].

## 4 Evaluation Metrics

Having decided on the logical representation schema for the ground-truth, we can develop metrics to assess the performance of TOC parsing systems. Our ground-truth representation for a TOC is a hierarchical list of triples. This

How to Be Happy  
Getting in Touch with Your Creativity  
From Nine to Five  
Getting Ahead  
One Faith

(a) A centered list (or table with one column) of *descriptors*. No *locators*.

CHAPTER IX  
THE THEISM OF RĀMĀNUJA . . . . . 659

Introduction—The Purāṇas—Life—History and literature—  
Bhāṣara—Yāgyavalkya—The Prāmānya—Implications of  
Rāmānuja's theory of knowledge—God—The individual soul—  
Matter—Creation—Ethics and religion—Mokṣa—General esti-  
mate.

CHAPTER X  
THE ŚAIVA, THE ŚĀKTA, AND THE LATER VAIṢṆAVA THEISM 722

Śaiva Siddhānta—Literature—Metaphysics, ethics and religion  
The Pratyabhijñā system of Kashmir—Śāktism—The dualism  
of Madhva—Life and literature—Theory of knowledge—God—  
Soul—Nature—God and the world—Ethics and religion—  
General estimate—Nimbārka and Keṣava—Vallabha—Caitanya,  
Jīva Gosvāmi and Baladeva.

(b) Only chapter level *entries* have *locators*. Section *descriptors* are hyphen separated and in reading order. Each chapter *descriptor* includes a chapter number (center aligned), and a chapter title which is left aligned on the next line.

10 Notes to the Reader  
12 Preface: Prospectors  
of the Peculiar  
18 Robert Ripley: The Man  
and the Phenomenon  
34 A Delicate Balance  
50 Dislocationists, Contortionists,  
and Elastic People

(c) *Locator* precedes *descriptor* in reading order. A semantic constraint that chapter numbers do not, but page-numbers may, increment in jumps is necessary to infer that the numbers are *locators*.

Figure 7. A variety of styles in Tables of Contents.

		
Introduction 1	Chapter One <b>BASIC MATERIALS</b> Everything You'll Need to Get Started • Essential Tools • More About Paper and Brushes 2	Chapter Two <b>TECHNIQUE</b> Using Your Materials Effectively • Four Basic Ways to Apply Color • Washes • Experimental Methods 18

Figure 8. Portion of a TOC that extends further to the right. The visual structure is columnar, but not tabular. Each column represents a chapter, with chapter number, title, bullet separated list of sections, and chapter locator.

can be represented as a tree of ordered pairs: (*descriptor*, *locator*), while enforcing sibling ordering in the tree.

If the target representation of a TOC analysis algorithm is also the same (which we would recommend), then tree comparisons metrics (such as tree edit distance [3]) become natural candidates for evaluation metrics. A tree edit distance measures the minimum number of operations (such as insertion, deletion, or substitution of tree nodes or subtrees) necessary to transform the parse-result tree into the ground-truth tree. Additional patterns of tree alteration can also be considered with appropriate penalty. For example, misdetecting a level change from visual cues will result in children becoming siblings or vice-versa. This pattern of error can be penalized according to the ultimate goal of the application.

However, we also consider evaluation metrics that do not require that the analysis output be in our ground-truth form. Rather, based on our ground-truth, we can frame queries for which we know the answer. We can then count how frequently (or how much) the ground-truth response agrees with the response of a system under evaluation. This way of evaluating parsing was suggested for tables in [7]. We can examine this idea by way of example.

An important, but not the only, use of a TOC is as a navigational aid. One might consider that the ultimate goal of TOC parsing is to establish navigational hyperlinks from TOC entries to actual locations of content. For this goal, it is important to correctly interpret *locator* indicators such as page numbers, the associations between *descriptor* and *locator* elements. Therefore, a TOC analyzer does not necessarily have to produce output in our format, or even fully interpret the hierarchy in a TOC. A simple listing of (*descriptor*, *locator*) pairs serves the purpose, and they need not even be in reading order.

A natural family of queries for evaluating the navigational task is: `TOC.getLocator(Descriptor query)` which can be asked for every entry in the ground-truth, and the expected response is the locator associated

with the entry in the ground-truth, or null if there is none. For each ground-truth entry.descriptor, let  $l$  denote the ground-truth *locator*, and  $l^*$  be the one returned by the test system.

We can compute a locator assignment recall metric as the average of the ratio:  $d(l, l^*)/d(l, l)$  computed over all *descriptor* items in ground-truth. Here  $d()$  denotes a similarity function defined for a pair of *locators*. Implicitly, of course, a comparison for equality will also have to be defined for *descriptors* to allow the system to find the best match to the query *descriptor*.

We could also turn this around. Let us say the goal is to associate with every *locator* a sufficiently descriptive chunk of the *descriptor*, while allowing for some segmentation error.

In a variant of the task, the correct ordering of *entries* may be considered important as well, especially if the children do not have associated *locator*. Also, we may want to assign higher penalties for errors in the higher levels of hierarchy. The constraints on expected output from automated systems will thus progressively increase. However, in all the above cases, as long as the ground-truth is in the suggested form, it is possible to measure performance of automated parsers that do not produce directly comparable output.

## 5 Implications for TOC Recognition Algorithms

Although we do not propose TOC recognition algorithms here, we suggest that our analysis places useful constraints on the nature of recognition algorithms that will be able to parse the vast diversity of tables of contents that fill libraries, magazine racks, web pages, and graphic designers' dreams.

**Recognition target:** We reiterate that tabular spatial layout is only sometimes employed, and when it is, the meanings of rows and columns are not predictable in advance but must be inferred from the TOC page on an individual basis. The target output of a recognition algorithm is not to be a true table, but a hierarchical tree of (*descriptor, locator, children*) nodes.

**Limits of Templates and Fixed Rules:** Any top-down, template-based approach will be limited in power to the templates that have been compiled. If the TOC domain is confined, for example, to journals of known layout, then pages from this family may indeed be successfully parsed. But the devices of TOC organization employ multiple cues that can be combined combinatorially. As such, these elements are *productive* of an open-ended design space, in the same sense that linguistic grammars are productive of an open-ended set of valid sentences. It will be impossible to enumerate all possible TOC designs that will be encountered in the general case.

**Style-dependent meaning of cues:** The mapping between visual cues such as alignment, spacing, font style, font size, ellipsis, etc. and the functional roles of the entities they are associated with, is variable according to each TOC's design. As is common with signal interpretation problems, purely local evidence is ambiguous and can be resolved only with regard to the context of surrounding material, or indeed, the whole.

**Multiple-level inference:** The main challenge and appeal of visual parsing problems is that they have to address multiple levels of inference. First, algorithms will have to be capable of gathering, and merging evidence and decisions at multiple scales of granularity. Second, functional segmentation or grouping of visual elements is exponentially complex [11]. The grouping of pixels (or characters if the TOC is in PDF format) into semantically valid chunks, their functional role interpretation, and reconciliation of individual functional roles to the overall syntactic and semantic structure will have to proceed hand in hand, perhaps through iterative steps of evidence exchange.

**Multiple levels of knowledge :** Finally multiple levels of knowledge: visual layout, style, syntactic structure (hierarchical list), semantic (*locators* are numbers), and pragmatic (*locators* mostly follow *descriptors* in reading order) will be often useful and sometimes necessary. Knowledge of stylistic conventions will certainly be important: it is very common for nodes of the same type and level to adopt parallel layout and attribute structure; it is common although by no means universal for locator nodes to obey columnar alignment; it is common for parent nodes to employ attributes of greater emphasis (e.g. larger font, boldface, capitalization) than their children in the hierarchy. Semantic knowledge may be critical in some cases, for example that chapter numbers increment by one while *locator* nodes increment monotonically but by several pages (see Figure 7c.)

We suggest that the class of recognition algorithms that can leverage these principles will have the following characteristics:

1. They will explore multiple hypotheses.
2. They will employ modular organization of knowledge, and appropriately factorized representation of hypotheses (for example, into logical content and visual style.)
3. They will combine weak evidence from multiple sources and types, such as layout, textual, linguistic, and semantic.
4. To manage the combinatorially large space of hypothesis they will quickly and cleverly eliminate, from consideration, large parts of this space on the basis of tiered evidence.

The evolution of document recognition algorithms possessing these characteristics will of course have impact on all

## POETRY

HÅKAN SANDELL	"Poetry rejoices..." Translated from the Swedish by Bill Coyle <i>Your Hair of Snakes and Flowers</i> Translated from the Swedish by Bill Coyle
YANNIS RITSOS	<i>The Crane Dance</i> Translated from the Greek by David Harsent
ANNA AKHMATOVA	<i>In Memoriam, July 19, 1914</i> Translated from the Russian by Stephen Edgar

**Figure 9. A TOC for which the two authors arrived at different ground-truth structures. In one, the structure is a flat list of descriptors with the poet missing (to be inferred) in some. In the other, poems are listed as children of the poet.**

versions of the page layout analysis problem, far beyond recognition of tables of contents. To date, the table of contents recognition work of Bourgeois *et al.* [4] stands as closest to in spirit to these ideals, although it is limited to trained models for specific journals, with a fixed set of levels.

## 6 Conclusion and Prognosis

In this paper we have attempted to identify principles underlying the phenomena of tables of contents as an important class of document that, in the future, our field must master. The tools and machinery for table of contents recognition have yet to be developed. Yet we can begin to see the outlines for what this technology will entail.

Tables of contents are inventions of human cultures. When we thoroughly understand the elements of layout and attribute styles and their mappings to logical structure, to rhetorical devices, to cultural conventions, and to emotional intent, we will be able not only to transform scanned printed books into databases and to hyperlinked active pages, but we will be equipped to support investigations at a larger scope. For example, we might take a historical view to chart the introduction of stylistic devices through time and across cultural mixing. We might extrapolate to anticipate new forms for tables of contents which are foreign to today's readers but may become trends in the future. We might create new genres of documents that seamlessly provide information while customizing its organization to the reader's personal knowledge. We might construct texts tailored to children's individual visual developmental stages. In short, the possibilities for applying theories for document recognition are open-ended, as are the tables of contents we encounter today.

**Acknowledgment** We would like to acknowledge insightful discussions with Mark Stefik at the Palo Alto Research Center.

## References

- [1] M. Agrawala and C. Stolte. Rendering effective route maps: Improving usability through generalization. In *Proc. ACM SIGGRAPH*, Los Angeles, USA, 2001. ACM.
- [2] A. Belaïd. Recognition of table of contents for electronic library consulting. *IJDAR*, 4(1):35–45, 2001.
- [3] P. Bille. A survey on tree edit distance and related problems. *Theor. Comput. Sci.*, 337(1-3):217–239, 2005.
- [4] F. L. Bourgeois, H. Emptoz, and S. S. Bensafi. Document understanding using probabilistic relaxation: Application on tables of contents of periodicals. In *ICDAR '01: Proceedings of the Sixth International Conference on Document Analysis and Recognition*, page 508, Washington, DC, USA, 2001. IEEE Computer Society.
- [5] J.-P. Chanod, B. Chidlovskii, H. Déjean, O. Fambon, J. Fuselier, T. Jacquin, and J.-L. Meunier. *From Legacy Documents to XML: A Conversion Framework*, volume 3652/2005 of *Lecture Notes in Computer Science*, pages 92–103. Springer Berlin / Heidelberg, September 2005.
- [6] H. Déjean and J.-L. Meunier. Structuring documents according to their table of contents. In *DocEng '05: Proceedings of the 2005 ACM symposium on Document engineering*, pages 2–9, New York, NY, USA, 2005. ACM.
- [7] J. Hu, R. Kashi, D. P. Lopresti, and G. Wilfong. Evaluating the performance of table processing algorithms. *International Journal on Document Analysis and Recognition*, 4(3):140–153, March 2002.
- [8] X. Lin and Y. Xiong. Detection and analysis of table of contents based on content association. *International Journal on Document Analysis and Recognition*, 8(2-3):132–143, 2006.
- [9] S. Mandal, S. P. Chowdhury, A. K. Das, and B. Chanda. Automated detection and segmentation of table of contents page from document images. In *ICDAR '03: Proceedings of the Seventh International Conference on Document Analysis and Recognition*, page 398, Washington, DC, USA, 2003. IEEE Computer Society.
- [10] G. Nagy. What does a machine need to know to read a document? In *Proceedings of ISRI Symp. on Document Analysis and Information Retrieval*, pages 1–10, Las Vegas, NV, March 1992.
- [11] M. Shilman, P. Liang, and P. Viola. Learning non-generative grammatical models for document analysis. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 962–969, Washington, DC, USA, 2005. IEEE Computer Society.
- [12] S. Tsuruoka, C. Hirano, T. Yoshikawa, and T. Shinogi. Image-based structure analysis for a table of contents and conversion to xml documents. In *Workshop on Document Layout Interpretation and Its Application (DLIA 2001)*, Seattle, USA, 2001.