In this chapter we will work an example of a stepwise formalization of mathematical knowledge. This is the task of e.g. an editor of a mathematical textbook preparing it for web-based publication. We will use an informal, but rigorous text: a fragment of Bourbaki’s Algebra [Bou74], which we show in Figure 4.1. We will mark it up in four stages, discussing the relevant OMDoc elements and the design decisions in the OMDoc format as we go along. Even though the text was actually written prior to the availability of the \TeX/L\TeX system, we will take a L\TeX representation as the starting point of our markup experiment, since this is the prevalent source markup format in mathematics nowadays.

Section 4.1 discusses the minimal markup that is needed to turn an arbitrary document into a valid OMDoc document — albeit one, where the markup is worthless of course. It discusses the necessary XML infrastructure and adds some meta-data to be used e.g. for document retrieval or archiving purposes.

In Section 4.2 we mark up the top-level structure of the text and classify the paragraphs by their category as mathematical statements. This level of markup already allows us to annotate and extract some meta-data and would allow applications to slice the text into individual units, store it in databases like MBase (see Section 26.4), or the In2Math knowledge base [Dah01, BB01], or assemble the text slices into individualized books e.g. covering only a sub-topic of the original work. However, all of the text itself, still contains the \TeX markup for formulae, which is readable only by experienced humans, and is fixed in notation. Based on the segmentation and meta-data, suitable systems like the ACTIVEMATH system described in Section 26.8 can re-assemble the text in different orders.

In Section 4.3 we will map all mathematical objects in the text into OpenMath or Content-MathML objects. To do this, we have to decide which symbols we want to use for marking up the formulae, and how to structure the theories involved. This will not only give us the ability to generate specialized and user-adaptive notation for them (see Chapter 25), but
1. LAWS OF COMPOSITION

Definition 1. Let $E$ be a set. A mapping of $E \times E$ is called a law of composition on $E$. The value $f(x, y)$ of $f$ for an ordered pair $(x, y) \in E \times E$ is called the composition of $x$ and $y$ under this law. A set with a law of composition is called a magma.

The composition of $x$ and $y$ is usually denoted by writing $x$ and $y$ in a definite order and separating them by a characteristic symbol of the law in question (a symbol which it may be agreed to omit). Among the symbols most often used are $+$ and $\cdot$, the usual convention being to omit the latter if desired; with these symbols the composition of $x$ and $y$ is written respectively as $x + y$, $x \cdot y$ or $xy$.

A law denoted by the symbol $+$ is usually called addition (the composition $x + y$ being called the sum of $x$ and $y$) and we say that it is written additively; a law denoted by the symbol $\cdot$ is usually called multiplication (the composition $x \cdot y = xy$ being called the product for $x$ and $y$) and we say that it is written multiplicatively.

In the general arguments of paragraphs 1 to 3 of this chapter we shall generally use the symbols $\top$ and $\bot$ to denote arbitrary laws of composition.

Examples. (1) The mappings $(X, Y) \mapsto X \cup Y$ and $(X, Y) \mapsto X \cap Y$ are laws of composition on the set of subsets of a set $E$.
(2) On the set $\mathbb{N}$ of natural numbers addition, multiplication, and exponentiation are laws of composition (the compositions of $x \in \mathbb{N}$ and $y \in \mathbb{N}$ under these laws being denoted respectively by $x + y$, $x \cdot y$, or $xy$) (Set Theory, III, §3, no. 4).
(3) Let $E$ be a set; the mapping $(X, Y) \mapsto X \circ Y$ is a law of composition on the set of subsets of $E \times E$ (Set Theory, II, §5, no. 3, Definition 6); the mapping $(f, g) \mapsto f \circ g$ is a law of composition on the set of mappings from $E$ into $E$ (Set Theory, II, §5, no. 2).

Fig. 4.1. A fragment from Bourbaki’s algebra [Bou74]

also to copy and paste them to symbolic math software systems. Furthermore, an assembly into texts can now be guided by the semantic theory structure, not only by the mathematical text categories or meta-data.

Finally, in Section 4.4 we will fully formalize the mathematical knowledge. This involves a transformation of the mathematical vernacular in the statements into some logical formalism. The main benefit of this is that we can verify the mathematical contents in theorem proving environments like Nuprl [CAB+86], HOL [GM93], Mizar [Rud92] and OMEGA [BCF+97].

4.1 Minimal OMDoc Markup

It actually takes very little change to an existing document to make it a valid OMDoc document. We only need to wrap the text into the appropriate XML document tags. In Listing 4.2 we have done this and also added meta-data.