WIKIPEDIA

# Free monoid

In abstract algebra, the **free monoid** on a set is the monoid whose elements are all the finite sequences (or strings) of zero or more elements from that set, with string concatenation as the monoid operation and with the unique sequence of zero elements, often called the empty string and denoted by ε or λ, as the identity element. The free monoid on a set $A$ is usually denoted $A^*$. The **free semigroup** on $A$ is the subsemigroup of $A^*$ containing all elements except the empty string. It is usually denoted $A^+$.[1][2]

More generally, an abstract monoid (or semigroup) $S$ is described as **free** if it is isomorphic to the free monoid (or semigroup) on some set.[3]

As the name implies, free monoids and semigroups are those objects which satisfy the usual universal property defining free objects, in the respective categories of monoids and semigroups. It follows that every monoid (or semigroup) arises as a homomorphic image of a free monoid (or semigroup). The study of semigroups as images of free semigroups is called combinatorial semigroup theory.

## Contents



## Examples

### Natural numbers

The monoid ($\mathbf{N_0}$,+) of natural numbers (including zero) under addition is a free monoid on a singleton free generator, in this case the natural number 1. According to the formal definition, this monoid consists of all sequences like "1", "1+1", "1+1+1", "1+1+1+1", and so on, including the empty sequence. Mapping each such sequence to its evaluation result [4] and the empty sequence to zero establishes an isomorphism from the set of such sequences to $\mathbf{N_0}$. This isomorphism is compatible with "+", that is, for any two sequences $s$ and $t$, if $s$ is mapped (i.e. evaluated) to a number $m$ and $t$ to $n$, then their concatenation $s+t$ is mapped to the sum $m+n$.

### Kleene star

In formal language theory, usually a finite set of "symbols" A (sometimes called the alphabet) is considered. A finite sequence of symbols is called a "word over $A$", and the free monoid $A^*$ is called the "Kleene star of $A$". Thus, the abstract study of formal languages can be thought of as the study of subsets of finitely generated free monoids. There are deep connections between the theory of semigroups and that of automata. For example, the regular languages over $A$ are the homomorphic pre-images in $A^*$ of subsets of finite monoids.

For example, assuming an alphabet $A = \{a, b, c\}$, its Kleene star $A^*$ contains all concatenations of $a$, $b$, and $c$:

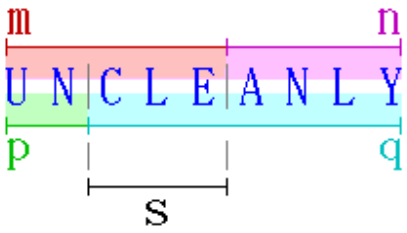
$\{\varepsilon, a, ab, ba, caa, cccbabbc, \ldots\}$.

If $A$ is any set, the *word length* function on $A^*$ is the unique monoid homomorphism from $A^*$ to ($\mathbf{N_0}$,+) that maps each element of $A$ to 1. A free monoid is thus a **graded monoid**.[5]

More generally, the regular languages over an alphabet A are the closure of the finite subsets of A*, the free monoid over A, under union, product, and generation of submonoid.[6]

## Conjugate words

We define a pair of words in $A^*$ of the form $uv$ and $vu$ as **conjugate**: the conjugates of a word are thus its circular shifts.[7] Two words are conjugate in this sense if they are conjugate in the sense of group theory as elements of the free group generated by $A$.[8]



Example for 1st case of equidivisibility: m="UNCLE", n="ANLY", p="UN", q="CLEANLY", and s="CLE"

### Equidivisibility

A free monoid is **equidivisible**: if the equation $mn = pq$ holds, then there exists an $s$ such that either $m = ps$, $sn = q$ (example see image) or $ms = p$, $n = sq$.[9] This result is also known as Levi's lemma.[10]

A monoid is free if and only if it is graded and equidivisible.[9]

## Free generators and rank

The members of a set $A$ are called the **free generators** for $A^*$ and $A^+$. The superscript * is then commonly understood to be the Kleene star. More generally, if $S$ is an abstract free monoid (semigroup), then a set of elements which maps onto the set of single-letter words under an isomorphism to a semigroup $A^+$ (monoid $A^*$) is called a *set of free generators* for $S$.

Each free semigroup (or monoid) $S$ has exactly one set of free generators, the cardinality of which is called the *rank* of $S$.

Two free monoids or semigroups are isomorphic if and only if they have the same rank. In fact, *every* set of generators for a free semigroup or monoid $S$ contains the free generators (see definition of generators in Monoid) since a free generator has word length 1 and hence can only be generated by itself. It follows that a free semigroup or monoid is finitely generated if and only if it has finite rank.

A submonoid $N$ of $A^*$ is **stable** if $u$, $v$, $ux$, $xv$ in $N$ together imply $x$ in $N$.[11] A submonoid of $A^*$ is stable if and only if it is free.[12] For example, using the set of bits { "0", "1" } as $A$, the set $N$ of all bit strings containing evenly many "1"s is a stable submonoid because if $u$ contains an even number of "1"s, and $ux$ as well, then $x$ must contain an even number of "1"s, too. While $N$ cannot be freely generated by any set of single bits, it *can* be freely generated by the set of bit strings { "0", "11", "101", "1001", "10001", ... }.

### Codes

A set of free generators for a free monoid $P$ is referred to as a **basis** for **P**: a set of words $C$ is a **code** if $C^*$ is a free monoid and $C$ is a basis.[3] A set $X$ of words in $A^*$ is a **prefix**, or has the **prefix property**, if it does not contain a proper (string) prefix of any of its elements. Every prefix in $A^+$ is a code, indeed a prefix code.[3][13]

A submonoid $N$ of $A^*$ is **right unitary** if $x$, $xy$ in $N$ implies $y$ in $N$. A submonoid is generated by a prefix if and only if it is right unitary.[14]

## Free hull

The intersection of free submonoids of a free monoid $A^*$ is again free.[15][16] If $S$ is a subset of a free monoid $A^*$ then the intersection of all free submonoids of $A^*$ containing $S$ is well-defined, since $A^*$ itself is free, and contains $S$; it is a free monoid and called the **free hull** of $S$. A basis for this intersection is a code.

The **defect theorem**[15][16][17] states that if $X$ is finite and $C$ is the basis of the free hull of $X$, then either $X$ is a code and $C = X$, or

$$|C| \leq |X| - 1 \ .$$

## Morphisms

A monoid morphism $f$ from a free monoid $B^*$ to a monoid $M$ is a map such that $f(xy) = f(x) \cdot f(y)$ for words $x, y$ and $f(\varepsilon) = \iota$, where $\varepsilon$ and $\iota$ denotes the identity element of $B^*$ and $M$, respectively. The morphism $f$ is determined by its values on the letters of $B$ and conversely any map from $B$ to $M$ extends to a morphism. A morphism is **non-erasing**[18] or **continuous**[19] if no letter of $B$ maps to $\iota$ and **trivial** if every letter of $B$ maps to $\iota$.[20]

A morphism $f$ from a free monoid $B^*$ to a free monoid $A^*$ is **total** if every letter of $A$ occurs in some word in the image of $f$; **cyclic**[20] or **periodic**[21] if the image of $f$ is contained in $\{w\}^*$ for some word $w$ of $A^*$. A morphism $f$ is ***k*-uniform** if the length $|f(a)|$ is constant and equal to $k$ for all $a$ in $A$.[22][23] A 1-uniform morphism is **strictly alphabetic**[19] or a **coding**.[24]

A morphism $f$ from a free monoid $B^*$ to a free monoid $A^*$ is **simplifiable** if there is an alphabet $C$ of cardinality less than that of $B$ such the morphism $f$ factors through $C^*$, that is, it is the composition of a morphism from $B^*$ to $C^*$ and a morphism from that to $A^*$; otherwise $f$ is **elementary**. The morphism $f$ is called a **code** if the image of the alphabet $B$ under $f$ is a code: every elementary morphism is a code.[25]

### Test sets

For $L$ a subset of $B^*$, a finite subset $T$ of $L$ is a *test set* for $L$ if morphisms $f$ and $g$ on $B^*$ agree on $L$ if and only if they agree on $T$. The **Ehrenfeucht conjecture** is that any subset $L$ has a test set:[26] it has been proved[27] independently by Albert and Lawrence; McNaughton; and Guba. The proofs rely on Hilbert's basis theorem.[28]

## Endomorphisms

An **endomorphism** of $A^*$ is a morphism from $A^*$ to itself.[29] The identity map $I$ is an endomorphism of $A^*$, and the endomorphisms form a monoid under composition of functions.

An endomorphism $f$ is **prolongable** if there is a letter $a$ such that $f(a) = as$ for a non-empty string $s$.[30]

### String projection

The operation of string projection is an endomorphism. That is, given a letter $a \in \Sigma$ and a string $s \in \Sigma^*$, the string projection $p_a(s)$ removes every occurrence of $a$ from $s$; it is formally defined by

$$p_a(s) = \begin{cases} \varepsilon & \text{if } s = \varepsilon, \text{ the empty string} \\ p_a(t) & \text{if } s = ta \\ p_a(t)b & \text{if } s = tb \text{ and } b \neq a. \end{cases}$$

Note that string projection is well-defined even if the rank of the monoid is infinite, as the above recursive definition works for all strings of finite length. String projection is a morphism in the category of free monoids, so that

$$p_a\left(\Sigma^*\right) = (\Sigma - a)^*$$

where $p_a\left(\Sigma^*\right)$ is understood to be the free monoid of all finite strings that don't contain the letter $a$. The identity morphism is $p_\varepsilon$, as clearly $p_\varepsilon(s) = s$ for all strings $s$. Of course, it commutes with the operation of string concatenation, so that $p_a(st) = p_a(s)p_a(t)$ for all strings $s$ and $t$. There are many right inverses to string projection, and thus it is a split epimorphism.

String projection is commutative, as clearly

$$p_a(p_b(s)) = p_b(p_a(s)).$$

For free monoids of finite rank, this follows from the fact that free monoids of the same rank are isomorphic, as projection reduces the rank of the monoid by one.

String projection is idempotent, as

$$p_a(p_a(s)) = p_a(s)$$

for all strings $s$. Thus, projection is an idempotent, commutative operation, and so it forms a bounded semilattice or a commutative band.

### Sturmian endomorphisms

An endomorphism of the free monoid $B^*$ on a 2-letter alphabet $B$ is **Sturmian** if it maps every Sturmian word to a Sturmian word[31][32] and **locally Sturmian** if it maps some Sturmian word to a Sturmian word.[33] The Sturmian endomorphisms form a submonoid of the monoid of endomorphisms of $B^*$.[31]

Define endomorphisms φ and ψ of $B^*$, where $B = \{0,1\}$, by $\varphi(0) = 01$, $\varphi(1) = 0$ and $\psi(0) = 10$, $\psi(1) = 0$. Then $I$, φ and ψ are Sturmian,[34] and the Sturmian endomorphisms of $B^*$ are precisely those endomorphisms in the submonoid of the endomorphism monoid generated by $\{I,\varphi,\psi\}$.[32][33][35]

A primitive substitution is Sturmian if the image of the word 10010010100101 is balanced.[32][36]

## The free commutative monoid

Given a set $A$, the **free commutative monoid** on $A$ is the set of all finite multisets with elements drawn from $A$, with the monoid operation being multiset sum and the monoid unit being the empty multiset.

For example, if $A = \{a, b, c\}$, elements of the free commutative monoid on $A$ are of the form

$$\{\varepsilon, a, ab, a^2b, ab^3c^4, \ldots\}.$$

The fundamental theorem of arithmetic states that the monoid of positive integers under multiplication is a free commutative monoid on an infinite set of generators, the prime numbers.

The **free commutative semigroup** is the subset of the free commutative monoid which contains all multisets with elements drawn from $A$ except the empty multiset.

## Generalization

The free partially commutative monoid, or *trace monoid*, is a generalization that encompasses both the free and free commutative monoids as instances. This generalization finds applications in combinatorics and in the study of parallelism in computer science.

## Free monoids and computing

The free monoid on a set $A$ corresponds to lists of elements from $A$ with concatenation as the binary operation. A monoid homomorphism from the free monoid to any other monoid $(M,\bullet)$ is a function $f$ such that

- $f(x_1...x_n) = f(x_1) \bullet ... \bullet f(x_n)$
- $f() = e$

where $e$ is the identity on $M$. Computationally, every such homomorphism corresponds to a map operation applying $f$ to all the elements of a list, followed by a fold operation which combines the results using the binary operator •. This computational paradigm (which can be generalised to non-associative binary operators) has inspired the MapReduce software framework.

# See also

- String operations

# Notes

1. Lothaire (1997, pp. 2–3), [1] (https://books.google.com/books?id=eATLTZzwW-sC&pg=PA2)
2. Pytheas Fogg (2002, p. 2)
3. Lothaire (1997, p. 5)
4. Since addition of natural numbers is associative, the result doesn't depend on the order of evaluation, thus ensuring the mapping to be well-defined.
5. Sakarovitch (2009) p.382
6. Borovik, Alexandre (2005-01-01). *Groups, Languages, Algorithms: AMS-ASL Joint Special Session on Interactions Between Logic, Group Theory, and Computer Science, January 16-19, 2003, Baltimore, Maryland* (https://books.google.com/books?id=C5QbCAAAQBAJ). American Mathematical Soc. ISBN 9780821836187.
7. Sakarovitch (2009) p.27
8. Pytheas Fogg (2002, p. 297)
9. Sakarovitch (2009) p.26
10. Aldo de Luca; Stefano Varricchio (1999). *Finiteness and Regularity in Semigroups and Formal Languages*. Springer Berlin Heidelberg. p. 2. ISBN 978-3-642-64150-3.
11. Berstel, Perrin & Reutenauer (2010, p. 61)
12. Berstel, Perrin & Reutenauer (2010, p. 62)
13. Berstel, Perrin & Reutenauer (2010, p. 58)
14. Lothaire (1997, p. 15)
15. Lothaire (1997, p. 6)
16. Lothaire (2011, p. 204)
17. Berstel, Perrin & Reutenauer (2010, p. 66)
18. Lothaire (1997, p. 7)
19. Sakarovitch (2009, p. 25)
20. Lothaire (1997, p. 164)
21. Salomaa (1981) p.77
22. Lothaire (2005, p. 522)
23. Berstel, Jean; Reutenauer, Christophe (2011). *Noncommutative rational series with applications*. Encyclopedia of Mathematics and Its Applications. **137**. Cambridge: Cambridge University Press. p. 103. ISBN 978-0-521-19022-0. Zbl 1250.68007 (https://zbmath.org/?format=complete&q=an:1250.68007).
24. Allouche & Shallit (2003, p. 9)
25. Salomaa (1981) p.72

26. Lothaire (1997, pp. 178–179)
27. Lothaire (2011, p. 451)
28. Salomaa, A. (October 1985). "The Ehrenfeucht conjecture: A proof for language theorists". *Bulletin of the EATCS* (27): 71–82.
29. Lothaire (2011, p. 450)
30. Allouche & Shallit (2003) p.10
31. Lothaire (2011, p. 83)
32. Pytheas Fogg (2002, p. 197)
33. Lothaire (2011, p. 85)
34. Lothaire (2011, p. 84)
35. Berstel, J.; Séébold, P. (1994). "A remark on morphic Sturmian words". *RAIRO, Inform. Théor. Appl. 2*. **8** (3–4): 255–263. doi:10.1051/ita/1994283-402551 (https://doi.org/10.1051%2Fita%2F1994283-402551). ISSN 0988-3754 (https://www.worldcat.org/issn/0988-3754). Zbl 0883.68104 (https://zbmath.org/?format=complete&q=an:0883.68104).
36. Berstel, Jean; Séébold, Patrice (1993), "A characterization of Sturmian morphisms", in Borzyszkowski, Andrzej M.; Sokołowski, Stefan (eds.), *Mathematical Foundations of Computer Science 1993. 18th International Symposium, MFCS'93 Gdańsk, Poland, August 30–September 3, 1993 Proceedings*, Lecture Notes in Computer Science, **711**, pp. 281–290, doi:10.1007/3-540-57182-5_20 (https://doi.org/10.1007%2F3-540-57182-5_20), ISBN 978-3-540-57182-7, Zbl 0925.11026 (https://zbmath.org/?format=complete&q=an:0925.11026)

# References


- Allouche, Jean-Paul; Shallit, Jeffrey (2003), *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, ISBN 978-0-521-82332-6, Zbl 1086.11015 (https://zbmath.org/?format=complete&q=an:1086.11015)
- Berstel, Jean; Perrin, Dominique; Reutenauer, Christophe (2010), *Codes and automata*, Encyclopedia of Mathematics and its Applications, **129**, Cambridge: Cambridge University Press, ISBN 978-0-521-88831-8, Zbl 1187.94001 (https://zbmath.org/?format=complete&q=an:1187.94001)
- Lothaire, M. (1997), *Combinatorics on words*, Cambridge Mathematical Library, **17**, Contributors: Perrin, D.; Reutenauer, C.; Berstel, J.; Pin, J. E.; Pirillo, G.; Foata, D.; Sakarovitch, J.; Simon, I.; Schützenberger, M. P.; Choffrut, C.; Cori, R. Series editors: Lyndon, Roger; Rota, Gian-Carlo. Foreword by Roger Lyndon (2nd ed.), Cambridge University Press, doi:10.1017/CBO9780511566097 (https://doi.org/10.1017%2FCBO9780511566097), ISBN 0-521-59924-5, MR 1475463 (https://www.ams.org/mathscinet-getitem?mr=1475463), Zbl 0874.20040 (https://zbmath.org/?format=complete&q=an:0874.20040)
- Lothaire, M. (2011), *Algebraic combinatorics on words*, Encyclopedia of Mathematics and Its Applications, **90**, With preface by Jean Berstel and Dominique Perrin (Reprint of the 2002 hardback ed.), Cambridge University Press, ISBN 978-0-521-18071-9, Zbl 1221.68183 (https://zbmath.org/?format=complete&q=an:1221.68183)
- Lothaire, M. (2005), *Applied combinatorics on words* (https://archive.org/details/appliedcombinato0000loth), Encyclopedia of Mathematics and Its Applications, **105**, A collective work by Jean Berstel, Dominique Perrin, Maxime Crochemore, Eric Laporte, Mehryar Mohri, Nadia Pisanti, Marie-France Sagot, Gesine Reinert, Sophie Schbath, Michael Waterman, Philippe Jacquet, Wojciech Szpankowski, Dominique Poulalhon, Gilles Schaeffer, Roman Kolpakov, Gregory Koucherov, Jean-Paul Allouche and Valérie Berthé, Cambridge: Cambridge University Press, ISBN 0-521-84802-4, Zbl 1133.68067 (https://zbmath.org/?format=complete&q=an:1133.68067)
- Pytheas Fogg, N. (2002), Berthé, Valérie; Ferenczi, Sébastien; Mauduit, Christian; Siegel, A. (eds.), *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, **1794**,

Berlin: Springer-Verlag, ISBN 3-540-44141-7, Zbl 1014.11015 (https://zbmath.org/?format=complete&q=an:1014.11015)

- Sakarovitch, Jacques (2009), *Elements of automata theory*, Translated from the French by Reuben Thomas, Cambridge: Cambridge University Press, ISBN 978-0-521-84425-3, Zbl 1188.68177 (https://zbmath.org/?format=complete&q=an:1188.68177)
- Salomaa, Arto (1981), *Jewels of Formal Language Theory*, Pitman Publishing, ISBN 0-273-08522-0, Zbl 0487.68064 (https://zbmath.org/?format=complete&q=an:0487.68064)

# External links

- Media related to Free monoid at Wikimedia Commons