La brève histoire de la théorie des automates et des langages formels

Dominique Perrin

- Introduction
- Automates finis
- Grammaires algébriques
- Calculabilité
- Séries formelles
- Combinatoire des mots

Les origines (1900-1940)

David Hilbert (1900) : Fondements des mathématiques et de la notion de preuve.

Axel Thue (1912) : Systèmes formels, combinatoire des mots.

Emil Post (1936), Modèles de calcul, problème de correspondance de Post.

Alonzo Church, Fonctions récursives, thèse de Church.

Alan Turing (1936), Calculabilité, machines de Turing

Les débuts (1940-60)

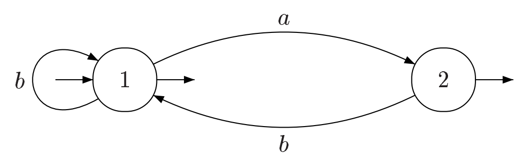
Claude Shannon (1949) : Théorie de l'information et de la communication.
Stephen Kleene (1956) : Automates finis
Noam Chomsky (1956) : Grammaires formelles

Hierarchie de Chomsky :

rationnel $\subset$ algébrique $\subset$ contextuel $\subset$ récursif

Automates finis

Warren McCulloch, W. Pitts (1943) : *A logical calculus of ideas immanent in nervous activity, Bull. Math. Biophisics.*



Stephen Kleene (1956) : *Representation of events in nerve nets and finite automata*, in *Automata Studies*, C. Shannon ed.

Expressions rationnelles

Opérations rationnelles (Kleene):

- union $X + Y$
- produit $XY = \{xy \mid x \in X, y \in Y\}$.
- étoile $X^* = \epsilon + X + \ldots$

Problèmes :

1. Identités
$$(a + b)^* = (a^*b)^*a^*$$
(théorie de John Conway):
2. Hauteur minimale en étoile?
Problème décidable : K. Hashiguchi, 1987.
3. Finite power property Brzozowski, 1964 : $X^* = \epsilon + X + \ldots + X$.

Le théorème de décomposition

Keneth Krohn, John Rhodes, 1965 : tout automate fini peut être obtenu en composant:

- des goupes de permutations
- des automates *reset* à 2 états.

Problème ouvert : complexité minimale

$$\mathcal{A} < \mathcal{A}_1 \circ \mathcal{G}_1 \circ \mathcal{A}_2 \circ \cdots \circ \mathcal{G}_n \circ \mathcal{A}_n$$

Semigroupes et variétés

Schützenberger, 1956 : Utilisation de semigroupes finis.

$$\varphi : A^* \to S \text{ morphisme de semigroupes}$$

Langage X reconnu par S si $X = \varphi^{-1}(S)$.

Variétés de semigroupes G. Birkhoff : famille fermée par morphismes, produits, sous-semigroupes.

McNaughton Langages sans étoile : produit+ toutes les opéations booléennes.

Théorème de Schützenberger (1965) : sans-étoile=apériodique.

Semigroupe apériodique : identité $x^n = x^{n+1}$

Automates et logique

Fragments décidables de la théorie des entiers naturels (Church).

Richard Büchi, 1962. Automates finis= logique monadique du second ordre des entiers avec le successeur.

Robert McNaughton, 1960, automates apériodiques= logique du premier ordre des entiers avec le successeur.

Objets infinis

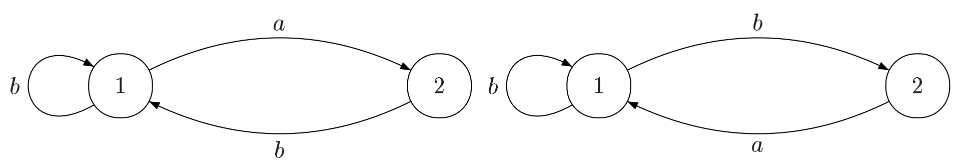
Richard Büchi, 1962 : automates sur les mots infinis.
Robert McNaughton, 1966 : déterminisation des automates sur les mots infinis (automates de Muller).
Michael Rabin, 1969 : les familles d'arbres infinis reconnaissables par automate fini sont fermées par complément.
Gurevitch, Harrington, 1982 : lien avec la théorie des jeux.

Dynamique symbolique

Marston Morse, Gustav Hedlund, 1936. Systemes dynamiques symboliques (*Recurrent geodesics on a surface of negative curvature*, Morse, 1921)

systèmes de type fini$\subset$ systèmes sofiques

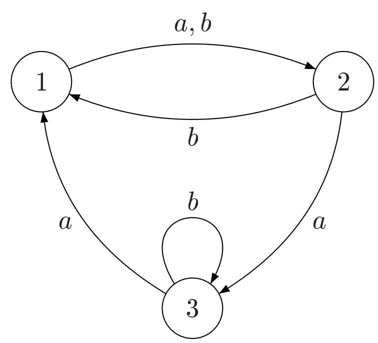
Problème ouvert : isomorphisme des systèmes de type fini.



Groupes et automates

J. Nielsen, 1921, Oscar Schreier, 1927 : Théorie combinatoire des groupes.

Algorithme de Todd-Coxeter (1936) : énumération des classes d'un sous-groupe(=calcul d'un automate fini).



William Thurston, 1992 : thérorie des groupes automatiques : présentation donnée par un automate fini→ groupes hyperboliques.

Langages context-free

Noam Chomsky, 1956 : Grammaires formelles.

$$\begin{aligned} phrase &\rightarrow sujet\ verbe\ complement \\ sujet &\rightarrow \cdots \end{aligned}$$

John Backus, 1959 : Syntaxe de ALGOL (Forme normale de Backus).

```
<expression> ::== <expression>+<expression>
<expression> ::== <expression>*<expression>
...
```

Analyse syntaxique et automates à pile

Analyse syntaxique ascendante et descendante.

Automate fini + mémoire en pile (premier entré/dernier sorti).

Donald Knuth, 1965 : analyse LR et LL.

Géraud Sénizergues, 1997 : décidabilité de l'équivalence des automates à pile déterministes (DPDA).

Calculabilité

Alan Turing, 1936, Machines de Turing : modèle mathématique d'un ordinateur de capacité non bornée.
Emil Post, 1947, Indécidabilité du problème des mots.

Complexité

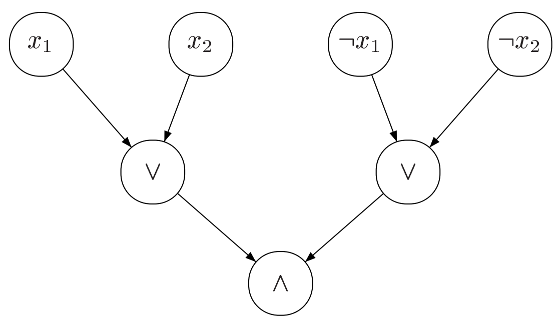
P = Classe des problèmes solubles en temps polynomial.
NP = Classe des problèmes solubles en temps polynomial de façon non déterministe= dont la solution est vérifiable en temps polynomial.
$PSPACE$ = Classe des problèmes solubles en espace polynomial.

$$P \subset NP \subset PSPACE$$

Problème ouvert : hierarchie stricte?

Circuits

Circuits calculant des fonctions booléennes de n variables $x_1, \cdots, x_n$ avec des portes ET et OU.
AC^0=classe des fonctions calculables par un circuit de taille polynomiale et hauteur constante.
NC^1=idem avec hauteur logarithmique mais degré entrant=2 .



Barrington,Thérien, 1987 : pour un langage rationnel AC^0=sans étoile.

Séries formelles

Schützenberger, 1962 : Séries formelles rationnelles et algébriques.
Exemple de série rationnelle : développement d'un entier en base 2

$$0 \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad 1 \rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

Exemple de série algébrique : énumération des arbres binaires á n noeuds

$$d(z) = \frac{1 - \sqrt{1 - 4z^2}}{2}$$

solution de l'équation $2z^2 d(z) - d(z) + 1 = 0$.

Combinatoire des mots

Problèmes combinatoires envisagés dès le début par Axel Thue, 1906.
Mot de Thue-Morse

$$t = abbabaab \cdots$$

puis

$$m = abcacbabcbac \cdots$$

est sans carré.

Mots Sturmiens

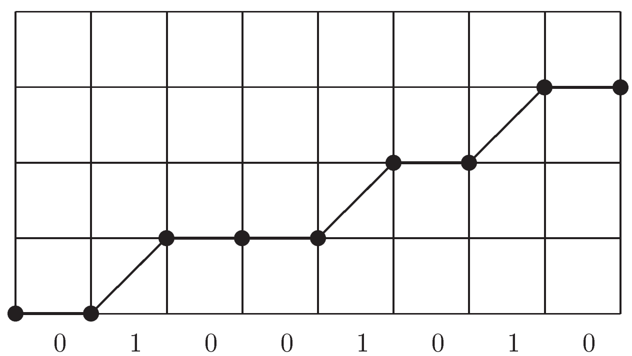
Morse, Hedlund, Mot de Fibonacci

$$f = 01001010\cdots$$

Interprétation graphique

$$s_n = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$$

avec $\alpha = 1/\varphi^2$ et $\varphi = (1+\sqrt{5})/2$.



Conclusion

Chapitre de l'informatique et des mathématiques lié à bien d'autres domaines.

- Logiciel (compilation, traduction, vérification)
- Communications (compression, codage)
- Algèbre appliquée (*computational algebra*)
- Langue naturelle (*computational linguistics*)
- Biologie (*computational biology*)