



# Recherche en Informatique

comment émergent les problèmes ?

Marie-Christine Rousset

Professeur



**!! Attention !!**

En raison d'un problème technique,  
l'enregistrement sonore ne commence  
qu'au transparent n°7 relatif à la  
Machine de Turing



L'informatique est une discipline très récente qui a vu le jour avec la construction des premiers ordinateurs dans la fin des années 30.

Depuis, elle a connu un essor considérable puisque l'informatique a pénétré tous les secteurs de la société, aussi bien dans la sphère professionnelle que privée. L'outil informatique, grâce à des ordinateurs de plus en plus petits et de plus en plus puissants, est devenu familier pour les professionnels de tous les secteurs mais aussi pour le grand public de tout âge. La science informatique est par contre mal connue et est souvent réduite à une simple mise au service d'autres sciences d'un outil puissant de calcul et de stockage de données.

Dans cet exposé, j'expliquerai d'abord ce qu'est la recherche en informatique, en montrant ses racines en mathématiques et son évolution au cours du temps. Je m'efforcerai ensuite d'illustrer au travers d'exemples différentes façons dont les problématiques de recherche émergent et se construisent en informatique.



# Recherche fondamentale

- Ancrée dans les Mathématiques
  - Modèles de calcul
    - Décidabilité et complexité de problèmes
      - Certains problèmes sont indécidables
    - Terminaison/correction/complétude d'algorithmes
    - Analyse de la complexité en temps et en place mémoire de l'exécution d'algorithmes
  - Modèles de données
    - Langages de requêtes « logiques » corrects et complets
  - Modèles de systèmes informatiques
    - Évaluation de performances (analytique ou expérimentale)
    - Vérification



# Les premiers problèmes

- Autour des systèmes formels et de la mécanisation du raisonnement
  - Un grand défi du début du 20e siècle en mathématiques
    - axiomatisation complète de l'arithmétique
      - un des 23 problèmes de Hilbert (1900)
    - trouver un système formel complet pour la logique des prédicats du premier ordre
      - tout énoncé vrai de la logique des prédicats du premier ordre peut-il être obtenu comme résultat d'un enchaînement de règles d'inférences à partir d'axiomes ?



# Résultats fondamentaux

- montrant les limites intrinsèques :

- des systèmes formels comme modèle universel de raisonnement :

- théorème(s) d'incomplétude de Gödel (1933)

pour toute théorie mathématique (englobant l'arithmétique), il existe des énoncés vrais dans cette théorie qui ne peuvent pas être prouvés à l'aide d'axiomes et de règles d'inférence d'un système formel

- de l'algorithmique et donc de l'informatique

- fondée sur le modèle de la Machine de Turing

- résultats d'indécidabilité de Turing (1936)

- Il existe des problèmes indécidables

- il n'existe pas d'algorithme permettant de les résoudre



# Machine de Turing

- Modèle abstrait d'un ordinateur
  - un ruban: séquence infinie de cellules (mémoire)
  - une tête de lecture qui peut lire et écrire un symbole dans la cellule où elle est positionnée et se déplacer d'une cellule vers la droite ou la gauche
    - un nombre fini de symboles et d'actions possibles
  - un nombre fini d'états
  
- Un algorithme
  - description des actions à effectuer en fonction de l'état et du symbole sous la tête de lecture (fonction de transition)
  - exécution d'un algorithme (calcul): application de la fonction de transition sur une entrée décrite par le contenu d'un nombre fini de cellules consécutives



# Universalité de ce modèle

- Thèse de Church-Turing:
  - ce modèle abstrait très simple capture tout ce que l'on peut calculer avec une machine
- Vérifiée dans la pratique:
  - tous les ordinateurs et les langages de programmation actuels sont Turing-complets
    - donc équivalents du point de vue de leur expressivité
- Fonde la théorie de la complexité en informatique
  - classes de complexité pour les problèmes décidables
    - en fonction des ressources (en temps, en place mémoire) nécessaires pour résoudre le problème
      - Temps: nombre d'étapes de calcul (en fonction de la taille de l'entrée)
      - Mémoire: nombre de cellules visitées (en fonction de la taille de l'entrée)

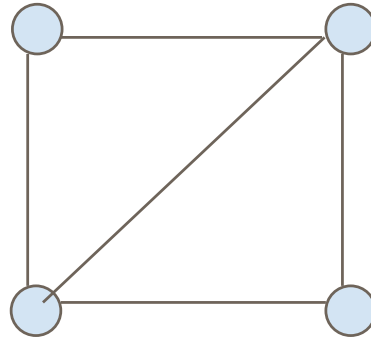


# Exemple: réseau de communication

## Modélisation :

### Graphe + Contraintes + Coûts

■ Graphe :



■ Contraintes : Aptitude des sommets/arcs à réaliser telle ou telle tâche

■ Coûts : mesure (temps, nombre de messages, etc.) de l'exécution de telle ou telle tâche

# Analyse de complexité

illustrée sur le problème de diffusion dans un graphe

temps-diffusion( $G,u$ ) = le “temps” nécessaire pour qu’une donnée soit transmise du sommet  $u$  à tous les sommets du graphe  $G$ .

Théorème: Décider si  $\text{temps-diffusion}(G,u) \leq k$  est **NP-complet**

- $P = \{ \text{problèmes résolubles en temps polynomial} \}$
- $NP = \{ \text{problèmes vérifiables en temps polynomial} \}$

**Conjecture :  $P \neq NP$**



# Comment contourner la complexité intrinsèque d'un problème ?

- Par des restrictions sur les instances du problème
- Par des algorithmes d'approximation

Algorithmes A

- de complexité polynômiale
- tels que, pour toute instance I du problème,

$$A(I) / \text{sol}(I) < r \text{ (rapport d'approximation)}$$

Théorème : Il existe un algorithme d'approximation pour le calcul du temps de diffusion avec  $r = O(\log n)$ , où n est la taille du graphe



# Recherche pluri-disciplinaire

## ■ Bioinformatique

- analyse et interprète au moyen de méthodes informatiques des données génomiques et leurs corrélations
- approche multi-critères
  - Combinatoire
  - Extraction de connaissances à partir de données
  - Visualisation de données
- en collaboration avec des biologistes



# Recherche parfois guidée par les avancées technologiques

## ■ Internet :

- retombée d'une recherche ciblée
  - besoin de l'armée américaine d'un réseau de communication capable de résister à une attaque nucléaire
    - Arpanet
  - émergence du principe d'une architecture décentralisée d'un réseau de réseaux

## ■ Web :

- retombée d'un outil de partage de documentations techniques pour les physiciens du CERN
- invention du langage html fondé sur le paradigme de la navigation entre documents par liens hypertextes



# Nouvelles voies de recherche

- ouvertes par ces avancées technologiques
  - nouveaux concepts
  - nouvelles problématiques
  - pour le passage à l'échelle de l'exploitation de l'immense potentiel de calcul et de stockage de données distribués sur le Web.



# Web sémantique

- Web : un immense gisement de données ou de services distribués et décrits par du texte
  - moteur de recherche par mots-clés
  - analyse statistique sommaire des textes et stockage de gros index de mots
- Enrichir la description des ressources par des méta-données ayant une sémantique formelle exploitable par des machines
  - construire des moteurs de requêtes plus puissants et plus précis que les moteurs de recherche actuel
  - en évitant les ambiguïtés des mots du langage naturel



# Calcul distribué à grande échelle

- Web: des ressources de calcul distribuées à grande échelle
  - global computing : SETI @home, Folding@home, XtremWeb
  - utilisation de centaines de milliers de PCs connectés sur Internet



# XtremWeb®

Connect your PC and  
participate to a world wide  
research experiment

*A global computing  
experimental platform*

XtremWeb est une plate-forme académique **pluridisciplinaire**  
pour l'étude des systèmes de Global Computing

L'objectif est de fournir aux chercheurs intéressés un support  
pour l'expérimentation

Nombreux problèmes informatiques posés :

- **Extensibilité** jusqu'à 1 million de machines
- **Hétérogénéité** : différents matériels et systèmes
- **Dynamisme**: le nombre de machines connectées évolue constamment
- **Disponibilité**: une ressource doit rester prioritairement disponible pour son propriétaire
- **Tolérance aux pannes**
- **Utilisable**: le système doit être programmable et maintenable
- **Sécurité**: pour les ressources, les serveurs, les applications

pour plus d'information : **[www.xtremweb.net](http://www.xtremweb.net)**



# Recherche aussi guidée par des conjectures scientifiques

- Algorithmique quantique : fondée sur un nouveau modèle de calcul issu de la mécanique quantique
- Nouvelle théorie de la complexité
  - certains problèmes (ex: factorisation de nombres) peuvent être résolus exponentiellement plus rapidement que dans le modèle de calcul de la Machine de Turing
    - problèmes dont la complexité exponentielle fonde la cryptographie
- Pari sur l'avenir
  - on n'est pas sûr qu'un jour un ordinateur quantique puisse être construit
  - mais si c'est possible révolutionnera l'informatique

# Digiteo Labs

