

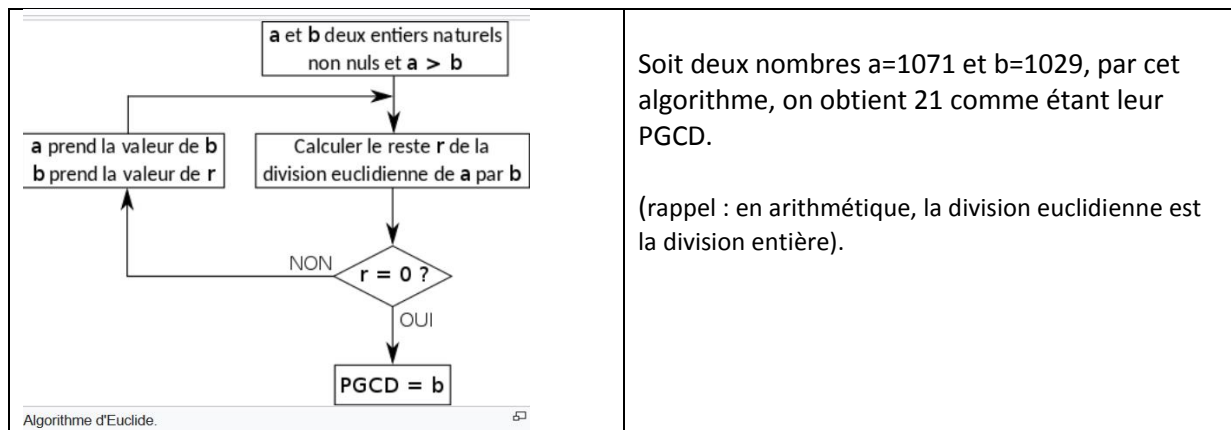
Atelier : Penser la technique, séance 1 : les algorithmes

Claude Poulain, 30/01/2008

Origine du terme

Muhammad Ibn Mūsā al-Khwarizmī, généralement appelé **Al-Khwarizmi**ⁿ¹ (latinisé en *Algoritmi*¹), né dans les années 780, originaire de Khiva dans la région du Khwarezm qui lui a donné son nom, dans l'actuel Ouzbékistanⁿ², mort vers 850 à Bagdad, est un mathématicien, géographe, astrologue et astronome perse^{2,3}, membre de la Maison de la sagesse de Bagdad. Ses écrits, rédigés en langue arabe, puis traduits en latin à partir du XII^e siècle, ont permis l'introduction de l'algèbre en Europe¹. Sa vie s'est déroulée en totalité à l'époque de la dynastie abbasside. (Wikipedia)

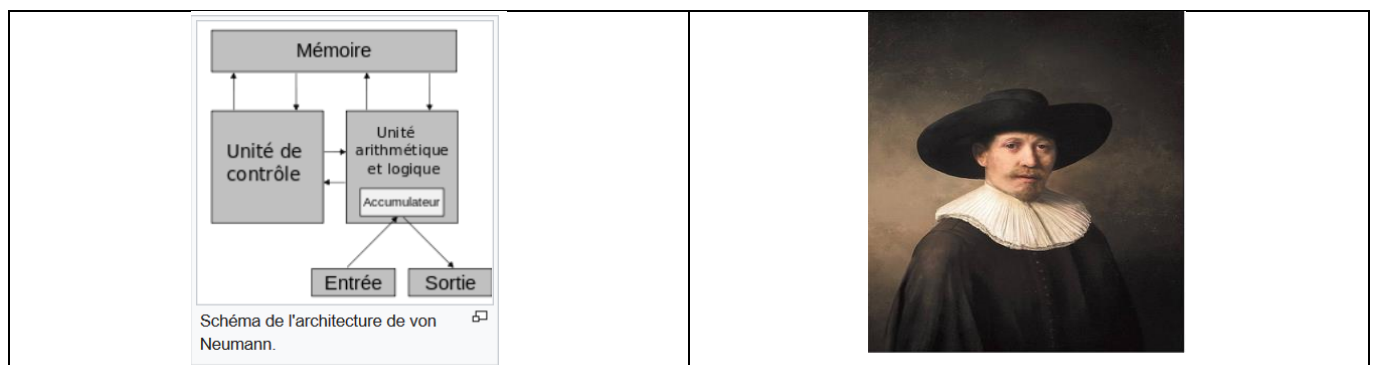
Algorithme d'Euclide (env -300) pour trouver le PGCD de deux nombres



Pourquoi parler des algorithmes en 2018 ?

-Développement du numérique : un seul système de représentation pour les nombres, les textes, les images, les sons, les ondes, etc. par une suite de signaux binaires. Ces signaux constituent l'information au sens de la théorie du même nom (Claude Shannon, 1916-2001).

-Développement des ordinateurs : machine qui fonctionne sur le même principe (la même architecture) depuis 1945. Seule a changé sa taille donc sa vitesse de calcul. Deux caractères importants : 1-Les données et le programme sont dans la même mémoire -> un programme peut donc se modifier lui-même. 2-C'est une machine universelle : un programme (texte d'instructions) va transformer le temps de son exécution une machine universelle, le calculateur, en une machine spécifique : machine à écrire, instrument de musique, calculatrice, décideur d'affectation des élèves, boule de cristal, producteur d'œuvre d'art (?) etc.



De l'administration des choses au gouvernement des hommes

Les grandes fonctions des algorithmes et de l'IA dans différents secteurs

	Education	Justice	Santé	Sécurité	Travail, RH	Culture	Autres
Générer de la connaissance	Mieux cerner les aptitudes d'apprentissage des élèves	Mettre en évidence les manières différenciées de rendre la justice selon les régions	Tirer profit de la quantité immense de publications scientifiques	Repérer des liens insoupçonnés pour la résolution d'enquêtes par les services de gendarmerie	Comprendre les phénomènes sociaux en entreprise	Créer des œuvres culturelles (peinture, musique)	Affiner le profil de risque d'un client d'un assureur
Faire du matching	Répartir les candidats au sein des formations d'enseignement supérieur (APB)		Répartir des patients pour participation à un essai clinique		Faire correspondre une liste de candidatures avec une offre d'emploi		Mettre en relation des profils « compatibles » sur des applications de rencontres, etc.
Prédire	Prédire des décrochages scolaires	Prédire la chance de succès d'un procès et le montant potentiel de dommages-intérêts	Prédire des épidémies Repérer des prédispositions à certaines pathologies afin d'en éviter le développement	Détecter les profils à risque dans la lutte anti-terroriste Prédire l'occurrence future de crimes et délits	Détecter les collaborateurs qui risquent de démissionner dans les prochains mois	Créer des œuvres ayant un maximum de chance de plaire aux spectateurs (Netflix)	
Recommander	Recommander des voies d'orientation personnalisées aux élèves	Recommander des solutions de médiation en fonction du profil des personnes et des cas similaires passés			Proposer des orientations de carrière adaptées aux profils des personnes	Recommander des livres (Amazon), des séries télévisées (Netflix), etc.	Individualiser des messages politiques sur les réseaux sociaux
Aider la décision		Suggérer au juge la solution jurisprudentielle la plus adéquate pour un cas donné	Suggérer au médecin des solutions thérapeutiques adaptées	Suggérer aux forces de police les zones prioritaires dans lesquelles patrouiller			Aider le conducteur à trouver le chemin le plus court d'un point à un autre (GPS)

Source : rapport CNIL, 2017, p.22

Comment parler des algorithmes ?

« Omniprésents, ces (algorithmes) restent pour nous mystérieux. Ils orientent les décisions, appaillent des processus automatiques et justifient des choix politiques, mais nous nous interrogeons rarement de la manière dont ils ont été produits. Nous regardons leurs effets sans examiner leur fabrication. Quelles sont les données qui servent au calcul ? Comment l'information a-t-elle été quantifiée ? Quels sont les principes de représentation qui animent le modèle statistique mis en œuvre pour classer tel objet plutôt que tel autre ? Qui pilote le codage des calculs et quels sont les objectifs ?

Habités par un sentiment d'incompétence, nous préférons ignorer les conditions de fonctionnement de la société des calculs, en laissant les clés aux statisticiens, aux informaticiens et aux économètres. La complexification des modèles algorithmiques mis en œuvre dans les nouvelles infrastructures informationnelles contribue à imposer le silence à ceux qui sont soumis à leurs effets. Elle désarme aussi ceux qui entreprennent de critiquer l'avènement de la froide rationalité des calculs, sans chercher à en comprendre le fonctionnement. Par facilité autant que par ignorance, la critique du nouvel empire des calculs s'est réfugiée dans une pseudo-opposition entre les « humains » et les « machines ». Elle dénonce confortablement la rationalisation néolibérale du monde, la tyrannie de l'évaluation ou les accidents automatisés du high-frequency trading.

Si cette défense constitue une sorte de contrepoison, elle reste plus gratifiante que véritablement efficace. La critique de la raison calculatoire ne peut opposer une rêverie pastorale à la marche automatisée des grands systèmes technologiques mondiaux. Pour vraiment critiquer une dynamique qui possède de si puissants moteurs économiques et culturels, il est nécessaire d'entrer dans les calculs, d'explorer leurs rouages et d'identifier leur vision du monde. Avant de réduire la logique calculatoire aux intérêts économiques de ceux qui la fabriquent, il faut commencer par allonger les algorithmes sur le divan et entendre la variété de leurs désirs ».

Dominique Cardon

« A quoi rêvent les algorithmes », Seuil, 2015, p.12-13

L'admission post-bac (APB)

Le problème à résoudre : Etant données quelques centaines de milliers de nouveaux bacheliers, il s'agit en quelques semaines de les affecter dans plus de 10.000 formations en respectant, autant que possible, leurs souhaits ainsi que des critères imposés par les formations d'accueil et/ou par la politique nationale, le tout dans la limite des places disponibles.

La solution mise en place :

- Un système centralisé : une base de données nationale à laquelle les lycéens d'une part, les formations d'autre part s'adressent via un site web.
- Chaque candidat dispose d'un « espace » personnel où figurent ses identifiants (INE, nom, prénom, adresse, e-mail, ...), son dossier scolaire, ses « vœux » (24 au maximum) classés selon son ordre de préférence, les réponses reçues. Du côté des formations, chacune à son espace où ranger les candidatures reçues, les réponses données, etc.
- Un calendrier unique rythme les dépôts de candidatures et leurs réponses (Il commence en janvier pour se terminer en septembre).
- L'affectation s'opère au moyen de deux méthodes distinctes, un **algorithme d'affectation** pour les formations dites « sélectives », un **système de classement** pour les autres.

En 1961 le ministère de l'enseignement a demandé à un professeur de Toulouse (Bernard Koehret, l'Institut National Polytechnique de Toulouse) de concevoir un système d'affectation des élèves dans les classes préparatoires aux grandes écoles (CPGE), ce qu'on dénomme les « prépas ». C'est pour cette population (environ 150.000 personnes) qu'il a conçu et mis en place dès 2004 un système utilisant un algorithme d'affectation appareillant d'une part les vœux des élèves et les critères des écoles. Cet algorithme est dérivé des recherches de deux mathématiciens, David Gale et Lloyd Shapley sur les « mariages stables ».

Mais la loi française considérant que le bac est un diplôme ouvrant à l'enseignement supérieur, la sélection à l'entrée de l'université est interdite. Observant des taux d'échec très élevés à la fin des premières années d'université, celles-ci (et beaucoup d'autres formations non considérées comme « grandes » écoles à commencer par les IUT), se sont battues pour devenir « sélectives », et pouvoir poser des critères d'admission... ce qui permettait de bénéficier de l'algorithme d'affectation mis au point pour les CPGE.

Pour les élèves qui candidatent à des formations qui sont restées « non sélectives », le système procède à un classement sur différents critères fixés par le ministère : 1/ les élèves provenant des lycées français de l'étranger, 2/les élèves de l'académie sont prioritaires par rapport à ceux des autres académies, 3/ le rang du vœux indiqué par l'élève, 4/un nombre aléatoire (qui sert au tirage au sort lorsque le nombre de candidats à la formation dépasse le nombre de places disponibles).

-

En 2017, à l'issue du premier tour, 644.000 élèves avaient une affectation mais 156.000 élèves n'en avaient pas. Aux derniers tours, les capacités d'accueil étant saturées, il a été procédé à un tirage au sort. A la fin des tours, il restait 10.000 élèves sans affectation.

Si le nombre d'élèves candidats augmente alors que le nombre de places disponibles n'augmente pas dans la même proportion, algorithme ou pas, il y aura un problème...

L'algorithme d'affectation d'APB

Rapport de la Cour des Comptes p.117-119

« L'algorithme d'affectation d'APB appartient à la famille des algorithmes de « mariage stable ». Ses fondements théoriques ont été établis en 1962 par les mathématiciens américains David Gale et Lloyd Shapley, dans un article fondateur qui s'attelait justement à résoudre le problème de l'admission à l'université. Cet algorithme a valu à ses auteurs, ainsi qu'au mathématicien Alvin Roth, ... le prix Nobel d'économie en 2012. Cette catégorie d'algorithme vise à résoudre des problèmes d'allocation sur des marchés que la mécanique du prix ne permet pas de résoudre.

Leurs travaux s'appuient sur la métaphore du marché du mariage, leur algorithme permet de marier des hommes et des femmes en vue de produire des mariages qui ont la propriété d'être stables parce qu'ils maximisent la satisfaction des deux parties et donc réduisent leur probabilité de rupture et d'appariement avec un autre partenaire.

Cet algorithme part du principe que la préférence absolue de chacun pour un seul partenaire est rarement réciproque. Elle est donc insuffisante pour que tous les hommes et femmes puissent contracter un mariage. Chacun doit donc exprimer des préférences de second rang. Il s'appuie ensuite sur un mécanisme d'acceptation différée.

La procédure d'appariement se déroule en plusieurs étapes :

- hommes et femmes établissent chacun une liste de conjoints potentiels, hiérarchisés et sans ex-aequo. ...l'intérêt rationnel de chacun est d'établir ce classement sincèrement, en fonction de ses préférences réelles, et non en supposant les préférences des autres ;*
- les hommes se proposent successivement aux femmes de leur liste de conjoints potentiels, dans l'ordre de préférences établi par eux ;*
- les femmes acceptent ou rejettent temporairement ces candidatures d'hommes en fonction de leur liste de préférences. Si la candidature est acceptée pour un temps, la femme s'autorise à rompre avec l'homme si, au tour suivant, un homme mieux placé dans sa liste de préférence se présente. Ce processus de proposition-acceptation-rejet temporaire se poursuit pour les hommes qui n'ont pas été retenus ou dont les « fiançailles » ont été rompues parce qu'un prétendant mieux classé est arrivé par la suite, jusqu'à ce que tous les hommes soient mariés.*

À la fin de cet algorithme d'acceptation différée, tous les mariages sont stables, c'est à dire qu'aucun homme et aucune femme n'a intérêt à quitter son conjoint pour un autre.... Si certaines personnes restent célibataires, il a été démontré qu'ils le seraient dans tous les modèles d'allocation.

*De façon générale, les algorithmes d'appariement, incluant de nombreuses modulations, trouvent beaucoup d'applications informatiques dans la sphère publique, en France comme à l'étranger. Pour le seul ministère de l'éducation nationale et de l'enseignement supérieur, on peut citer notamment : - **Galaxie** : ce portail du ministère de l'enseignement supérieur gère, entre autres, l'affectation des personnels enseignants;*

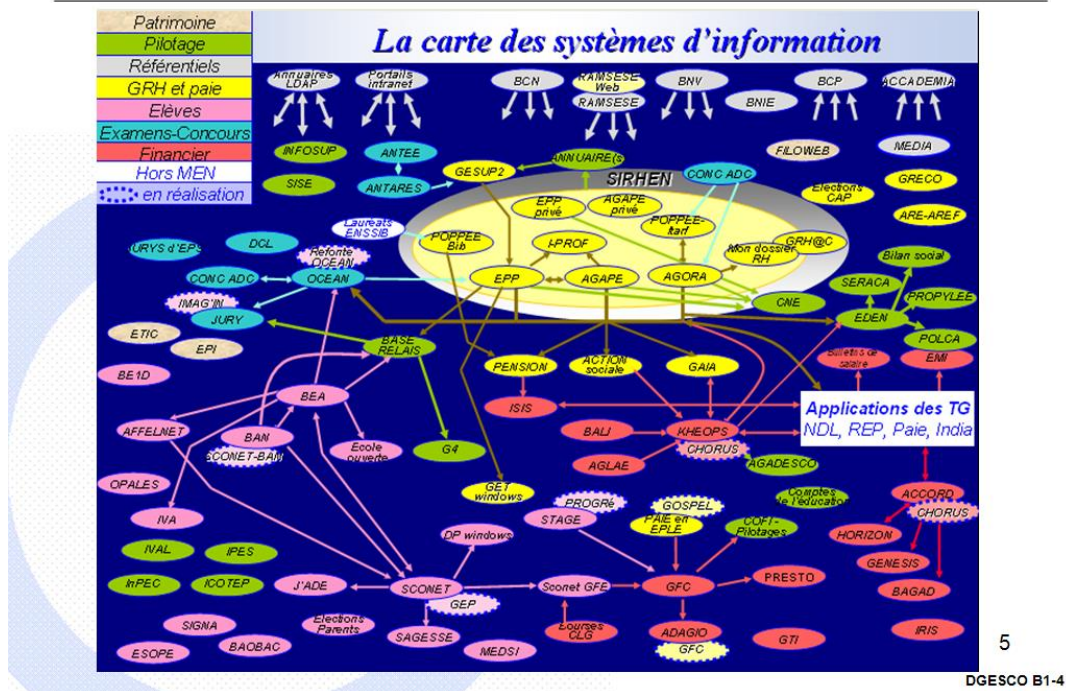
*- **Affelnet** : cette application gère depuis 2008 l'affectation des collégiens au lycée et, dans certaines académies, l'affectation des élèves de primaires au collège. - **le mouvement des enseignants du premier et du second degré** dans l'enseignement public ».*

Éléments d'analyse

L'algorithme APB est immergé dans un système de valeurs

- l'idéologie scolaire et, en France, la vénération du diplôme,
- les besoins de l'économie en personnes qualifiées : la formation scolaire et des diplômes sont encore des facteurs importants d'entrée dans l'emploi,
- l'idéologie gestionnaire en vigueur dans les grandes entreprises,
 - o d'où un système d'information qui n'a pas d'équivalent en France par sa taille et son détail.

Les SI



- 1- un fichage de tous les élèves, au départ en infraction à la loi informatique et liberté (pas de déclaration, variables personnelles collectées sans finalité, droit d'accès impossible, ...),
- 2- une ouverture problématique vers les mairies, les éditeurs de logiciels de « gestion » des établissements, annonce d'un accès donné à Microsoft.
- 3- Donne aux élèves l'habitude de la surveillance, de l'évaluation permanente, du calcul..
- 4- Il est de plus en plus difficile de vivre en dehors du système.

Faut-il faire confiance à l'algorithme d'APB ?

- Personne n'a jamais vu le texte de l'algorithme. C'est son auteur qui affirme qu'il est de la famille Gale-Shapley, et les experts confirment, au vu des « entrées » et des « sorties » que c'est possible.
- A supposer qu'on ait accès au texte, qui est capable de le lire et de le comprendre ?
- A supposer que quelqu'un en soit capable, qui peut assurer que c'est bien celui-là qui « tourne » dans l'ordinateur d'APB ?

Recommandation de la CNIL¹

« Rendre les systèmes algorithmiques compréhensibles en renforçant les droits existants et en organisant la médiation avec les utilisateurs ». Il faut que les auteurs soient joignables et qu'ils puissent expliquer. De plus, la CNIL recommande créer un service d'audit de ces algorithmes, de vérifier que l'on peut avoir des explications et que celles-ci sont loyales.

Le pouvoir absolu de l'algorithme

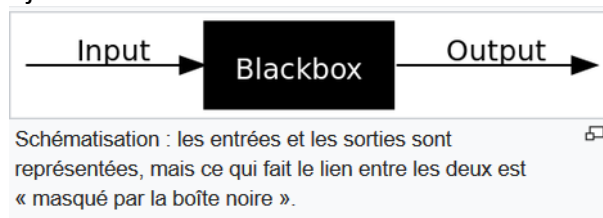
- APB donne une réponse sans aucune explication sur les traitements qui président à l'affectation comme aux refus de ses autres vœux. (Les professeurs n'en savent pas plus que leurs élèves).
- les élèves comme leurs professeurs ne savent pas c'est quels sont les critères mis en place par les formations², les informations utilisées, le poids de chacune, etc.

Cette pratique est illégale : Loi Informatique et libertés de 1978, Article 10 : « Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de sa personnalité ».

En 2017, la CNIL met en demeure le MEN de « cesser de prendre des décisions ... sur le seul fondement d'un traitement automatisé de données (ici APB) et de prévoir une intervention humaine permettant de tenir compte des observations des personnes ³ » ;

« La CNIL met surtout en demeure le MEN de mettre en œuvre une procédure effective de traitement des demandes de droit d'accès garantissant la transmission aux personnes des informations permettant de connaître et de contester la logique qui sous-tend le traitement APB ».

Pourquoi pas un algorithme « jouable » ?



A supposer que ce soit possible (ce l'est très probablement), qu'en pensez-vous ?

¹ Rapport CNIL, p.56

² Elles sont secrètes, décidées par chaque conseil d'administration.

³ décision n° 2017-053 du 30 août 2017 de la Commission nationale de l'informatique et des libertés mettant en demeure le ministère de l'enseignement supérieur, de la recherche et de l'innovation.

Incertitudes sur les critères et donc sur les données utilisées

- Quels sont les critères pour être soumis à APB ou non : quelle est la population qui sera soumise à l'algorithme ?
- Quels critères utiliser, quel degré de spécificité entre les formations ?
- Quelles données à utiliser pour « représenter » chacun des critères ?

Si on imagine que les notes et mentions au bac sont prises en compte, on est dans la plus grande imprécision quand ce n'est pas de l'incertitude pour tout le reste. Pourtant, une bonne part des inégalités scolaires se jouent dans ces critères.

« [Le système APB fait partie] de ces algorithmes qui véhiculent un processus de dépolitisation et de neutralisation de choix de société méritant pourtant de faire l'objet d'une discussion publique »⁴

Petite bibliographie

-**sur la théorie** : « Les algorithmes d'affectation des élèves aux établissements d'enseignement », Julien Grenet (CNRS et Ecole d'économie de Paris), Séminaire Codes Sources du Laboratoire Informatique de Paris 6, 2017 <https://www.youtube.com/watch?v=oprFPrZcNq4>.

-sur APB : Cour des Comptes, « ADMISSION POST-BAC ET ACCÈS À L'ENSEIGNEMENT SUPÉRIEUR, Un dispositif contesté à réformer, 2017.

-**sur l'analyse critique des algorithmes** : Commission Nationale Informatique et Libertés (CNIL), « Les enjeux éthiques des algorithmes et de l'intelligence artificielle », décembre 2017.

⁴ Rapport CNIL, p.22