



Ressources pour le cycle terminal général et technologique

Informatique et Sciences du Numérique

Vous avez dit trier ? 2 - Les critères de tri

Ces documents peuvent être utilisés et modifiés librement dans le cadre des activités d'enseignement scolaire, hors exploitation commerciale.

Toute reproduction totale ou partielle à d'autres fins est soumise à une autorisation préalable du Directeur général de l'enseignement scolaire.

La violation de ces dispositions est passible des sanctions édictées à l'article L.335-2 du Code de la propriété intellectuelle.

Juin 2012

Présentation / Vous avez dit trier ? 2 - Les critères de tri

1 / Contexte

La notion de tri, au-delà de l'aspect algorithmique évident et incontournable, pose la question du **critère**, qui est abordée dans ce second volet.

Pour les systèmes numériques, ce critère se traduit par un nombre (ou doit se résumer au final à un nombre), le tri d'un ensemble d'articles se résumant alors à ordonner un ensemble de nombres ; le problème est donc en relation directe avec les notions de numérisation et de codage. En revanche, les données non numériques ou soumises à des critères multiples obligent à mettre au point des critères plus complexes : tri des couleurs (voir ci-dessous), des fichiers dans un dossier, des réponses à une requête dans un moteur de recherche sur le Web, etc. Le choix des critères influence alors directement à la fois sur la performance du tri et sur l'usage qu'on va faire des données triées.



Par ailleurs, autour de cette question du critère de tri, peuvent éventuellement se cristalliser des éléments de débat et de travail pluridisciplinaire, impliquant fortement les sciences humaines, sur des thèmes comme les discriminations, l'eugénisme, l'égalité des chances, la mixité, etc.

Cette fiche propose un **scénario pédagogique** pour aborder quelques-uns de ces aspects.

2 / Thème abordé

2.1 Problématique

Comment ranger des données non numériques ?

Par essence, l'ordinateur ne sait trier que des nombres. Le tri d'une autre sorte d'objets suppose un codage numérique de certains paramètres de ces objets. Les possibilités de tri sont alors en rapport étroit avec le choix du code.

Par exemple, les différentes façons de coder la couleur¹ ne permettent pas les mêmes tris.

Quel est le meilleur coup dans un jeu de stratégie ?

L'objectif est ici de choisir, de décider quel doit être le prochain coup pour s'assurer la plus grande chance de victoire.

Au-delà des aspects tactiques, qui pourront d'ailleurs être abordés dans d'autres études algorithmiques (arborescence des coups, stratégie « minimax », etc.), l'attribution d'une note à chaque coup possible, afin d'en déterminer le meilleur, repose sur un ensemble de critères constituant la fonction d'évaluation d'une position.

Quel est le meilleur candidat pour ce poste ?

Là encore, l'objectif est de choisir, de décider. C'est un choix multicritères qui nécessite la mise en place d'une fonction d'évaluation.

Travailler sur de l'humain implique ici des aspects sociétaux et juridiques particuliers (voir les ressources).

Comment ordonnancer des données pour mettre en avant une propriété particulière ?

L'objectif est ici d'informer, de clarifier, d'expliquer, éventuellement de convaincre, voire de désinformer.

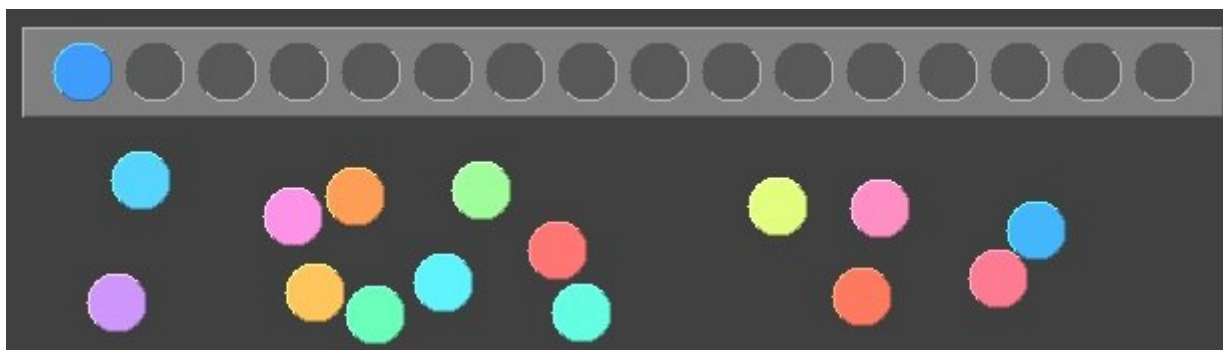
Le choix du ou des critères de classement d'un ensemble d'articles est porteur de sens ; par exemple, on peut trier les magasins d'une chaîne de distribution par résultat net, ou par chiffre d'affaire, ou par progression annuelle, et on ne voit pas la même chose. Et quand on trie des données selon un critère provoquant de nombreux « ex-æquo », on modifie souvent le critère pour départager les ex-æquo.

Autant de problématiques qui permettent d'introduire la notion de critère de tri, puis d'aborder les aspects sociétaux et juridiques induits.

2.2 Situation d'accroche

¹ Il s'agit essentiellement des systèmes RVB (Rouge, Vert, Bleu), CMJN (Cyan, Magenta, Jaune, Noir) et TLS (teinte, Luminosité, Saturation) : à chaque fois, trois ou quatre paramètres numériques influant sur le tri.

Pour déceler chez un patient une éventuelle anomalie de perception des couleurs (daltonisme), le personnel médical dispose d'un certain nombre de tests. Le test de Farnsworth consiste à ranger un ensemble de pions colorés de telle façon que des pions voisins portent des couleurs les plus « proches » possibles.



Comment un ordinateur pourrait-il s'y prendre pour réaliser cette opération de tri ?

Expérimentation, sous forme de TD par petits groupes.

Consignes :

- Pour commencer, faites le test de Farnsworth.
- Faites une copie d'écran du test terminé, puis ouvrez-la avec un programme de dessin (Paint, GIMP, etc.).
- Pour chaque pion, utilisez la pipette pour prélever la couleur, puis sélectionnez dans le menu « couleurs/modifier les couleurs/définir les couleurs personnalisées », et observez comment la couleur est codée dans l'ordinateur.
- Rendez compte de toutes vos observations par écrit.
- Une recherche documentaire sur les systèmes de représentation des couleurs peut aider.

2.3 Frontières de l'étude et prolongements possibles

Toute la problématique du critère de tri pourra être soulevée ; le tri multi-critères, en particulier, est intéressant si on trie des données organisées en colonnes (numériques ou textuelles) selon l'une des colonnes (par exemple, les prénoms) puis selon une autre colonne (par exemple, les noms de famille), est-ce que les ex-æquo sont « bien classés » (est-ce que DUPONT, Alain sera avant DUPONT, Antoine) ? C'est le problème de la **stabilité** des algorithmes de tri qui est posé ici.

Sur le plan humain, on peut y intégrer facilement des aspects sociétaux et juridiques, notamment sur le plan de la légitimité du tri et des critères employés pour sélectionner des personnes.

3 / Objectifs pédagogiques

3.1 Disciplines impliquées

Les aspects sociétaux et juridiques peuvent conduire à faire appel à des enseignants en sciences humaines ou en philosophie.

3.2 Prérequis

Aucun, mais il est utile d'avoir abordé auparavant quelques algorithmes (simples) de tri de nombres entiers afin que la problématique du tri soit déjà présente à l'esprit des élèves.

3.3 Éléments du programme

Contenus

Représentation de l'information :

- Numérisation : coder la couleur.
- Structuration et organisation de l'information : classer des informations.

Algorithmes :

- Algorithmes classiques, algorithmes de tri.

Compétences et capacités

Concevoir et réaliser une solution informatique en réponse à un problème :

- Comprendre et mettre en œuvre la notion de critère de tri.
- Comprendre et mettre en œuvre la notion de fonction d'évaluation.
- Imaginer des solutions à un problème complexe.

Collaborer efficacement au sein d'une équipe dans le cadre d'un projet :

- Participer à un groupe de discussion dans le respect des autres.

Faire un usage responsable des sciences du numérique :

- Avoir conscience de l'impact du numérique dans la société, de l'importance des licences et du droit.

4 / Modalités de mise en œuvre

4.1 Durée prévue

Entre deux et quatre heures, selon activités.

4.2 Type de l'animation

Petits groupes et classe entière, en alternance.

4.3 Éléments de cours / TP / TD

- Réflexion en classe entière (brain-storming) sur la notion de **critère de tri** et de **clé de tri** (dans le jeu de cartes, faut-il trier d'abord par valeurs ou d'abord par couleurs ? Quelle couleur mettre en « premier » ? que faire des atouts ?).
- Résolution de l'applet « test de Farnsworth » (TD).
- Réflexion en classe entière (brain-storming) sur la notion de critère de tri dans un **problème a priori non numérique** (test de Farnsworth). Nécessité de numériser le ou les critères pour le rendre le problème appréhendable par un système informatique.
- **Pertinence du codage** par rapport au tri envisagé (comparaison des codages RVB et HSL - des couleurs perçues comme proches sont-elles représentées par des nombres ou des groupes de nombres peu différents) (activité de type TD).
- Problématique de la **recherche du meilleur coup** dans un jeu de stratégie : il faut évaluer chacun des coups jouables dans une position donnée afin de pouvoir choisir le meilleur.²
Démarche active : comment attribuer une note à chacun des coups jouables d'une position donnée du jeu de morpion 3 x 3 (choix des critères, fonction d'évaluation).
Remarque : seule la fonction d'évaluation est étudiée ici, pas l'algorithme « minimax » (qui pourrait par ailleurs faire l'objet d'une autre étude).
- Possibilité de débat sur les **aspects sociétaux** (avec participation d'un enseignant de sciences humaines).

4.4 Recherche documentaire

On peut envisager quelques recherches sur le codage numérique de la couleur.

5 / Ressources, outils et références

Applet « test de Farnsworth » (tri de couleurs) :

<http://www.univie.ac.at/Vergl-Physiologie/colortest/colortestF-en.html>

Aspects sociétaux et juridiques :

- **Loi Pleven du 1er juillet 1972**
Toute discrimination, appel à la haine ou violence à l'égard d'une personne en raison de son origine, ou de son appartenance à une nation ou une religion seront punis d'une peine d'emprisonnement allant de 1 mois à 1 an et d'une amende comprise entre 300€ et 45000€.
- Déclaration universelle des droits de l'Homme, article 2
- Déclaration des droits de l'Homme et du Citoyen, article 1
- **Saïd Bouamama, Yvon Fotia et Jessy Cormont, « Les discriminations systémiques », in, Les Figures**

2 On a ici, à l'arrière-plan, l'**algorithme du plus court chemin**. La méthode donne une note à chaque suite de coups (ou stratégie) assimilée à un chemin dans un graphe.

de la Domination [En ligne], mis en ligne le : 27/10/2010, URL :
<http://www.lesfiguresdeladomination.org/index.php?id=542>

- « **Bienvenue à Gattaca** », film d'anticipation américain réalisé par Andrew Nicol, voir fiche
http://fr.wikipedia.org/wiki/Bienvenue_%C3%A0_Gattaca
- **Nathalie Georges**, « **Le profilage : outils statistique ou mode de coordination?** », Centre d'études de l'emploi, ENPC, novembre 2006, [En ligne] URL :
http://www.cee-recherche.fr/fr/doctrav/profilage_statistique_coordination_72.pdf

6 / Auteur

Jean-Paul Berthelot, groupe ISN, académie de Nantes.

Annexe – Exemple de fonction d'évaluation

1 / Contexte

Le sujet est le jeu de morpion 3 x 3.

Un programme de jeu de stratégie fonctionne ordinairement de la façon suivante (très schématiquement) :

- Il construit une arborescence, à partir de la position de jeu actuelle, en tenant compte de tous les coups qu'il peut jouer, puis de toutes les répliques possibles de l'adversaire pour chaque coup, et ainsi de suite.
- Comme il n'est en général pas possible de mener cette construction à son terme (pour des raisons de croissance exponentielle du nombre de cas à envisager), un algorithme est chargé d'élaguer les branches les moins intéressantes de l'arbre.
Même ainsi, la profondeur d'analyse autorisée par le temps de calcul ne permet en général pas d'atteindre la fin de la partie pour déterminer qui gagne.
- Il est donc nécessaire de disposer d'un moyen d'évaluer les positions obtenues au niveau des bourgeons de l'arbre, à des fins d'élagage d'une part, et à des fins d'évaluation de chacun des coups jouables dans la position actuelle.

Le procédé de calcul servant à attribuer une valeur numérique, une note, à une position du jeu s'appelle une fonction d'évaluation.

Ce procédé se déroule avec comme unique donnée la position à évaluer, sans souci de prévision des positions futures (c'est le travail de l'algorithme qui construit l'arborescence).

Lorsque le jeu est particulièrement simple, comme ici le morpion, il se peut qu'une fonction d'évaluation bien choisie soit suffisante pour trouver directement le ou les meilleurs coups. La construction de l'arborescence est alors inutile.

2 / Détermination d'une fonction d'évaluation pour le morpion 3 x 3

2.1 Quelques critères permettant de cerner un peu le problème

- Pour gagner, il est *a priori* préférable de jouer sur des lignes/colonnes/diagonales qui contiennent déjà un ou des pions de son propre camp, et pas encore de pion de l'adversaire (sinon ça ne sert à rien).
- Pour ne pas perdre, il est *a priori* préférable de jouer sur les lignes/colonnes/diagonales qui contiennent un ou des pions de l'adversaire, et pas encore de pion de son propre camp (sinon ça ne sert à rien).
- De ce qui précède, on peut conclure que les lignes/colonnes/diagonales qui contiennent au moins un pion de chaque camp sont « mortes » et n'ont plus d'intérêt dans la partie.
- À l'échelle d'un seul mouvement, « gagner » est plus important que « ne pas perdre », car cela entraîne la fin de la partie. Un coup gagnant tout de suite doit être joué, même si ce coup permet ensuite à l'adversaire d'aligner, car il n'en aura pas l'occasion, la partie étant terminée.

La fonction d'évaluation reçoit en entrée la position actuelle et le coup qu'il faut évaluer ; ce coup correspond à une case encore vide de la position.

2.2 Approche possible

Initialiser à zéro la note de la case à évaluer

Pour chaque alignement (ligne colonne diagonale) auquel appartient la case à évaluer :

- Compter les pions amis et les pions ennemis déjà présents sur cet alignement
- Ajuster la note de la case à évaluer selon la table (voir une proposition de table page suivante)

Fin pour chaque alignement

Exemple de table :

		Pions amis		
		0	1	2
Pions ennemis	0	+1	+30	+1000
	1	+10	0	0
	2	+100	0	0

Bien sûr, il est possible de discuter sur les valeurs à donner aux coefficients.

Une autre approche pourrait être de remplacer la table par une fonction mathématico-logique ayant pour arguments les nombres de pions amis et ennemis.

Par exemple, on pourra essayer :

$$f(\text{amis}, \text{ennemis}) = \text{abs}((\text{ennemis} = 0) * 30^{\text{amis}} + (\text{amis} = 0) * 10^{\text{ennemis}})$$

2.3 Remarque

La programmation est alors plus simple que le recours à une table ; cependant, le temps de traitement n'est pas forcément plus court (notamment ici, avec la présence de deux fonctions puissance dans le calcul).

Cela est sans conséquence pour un jeu simple comme le morpion. En revanche, pour un jeu plus complexe, comme la fonction d'évaluation a vocation à être exécutée un grand nombre de fois (autant que de positions explorées dans l'arbre), il est crucial d'en optimiser le temps de calcul car cela permettra, pour le même temps de « réflexion », d'explorer l'arbre plus profondément.

On comprend qu'il y a alors un compromis à trouver entre une fonction d'évaluation très complexe empêchant une exploration très profonde à cause du temps de calcul, et une fonction simple et rapide permettant d'explorer plus de positions (voie dite de la « force brute »).