



Vers la génération automatique de tests d'évaluations différenciés et équitables en contexte universitaire

Richardson CIGUENE, Céline JOIRON, Gilles DEQUEN

Atelier EIAEI 2015 – Mardi 2 juin 2015 – Agadir, Maroc



Contexte et objectifs

- Génération automatique de sujets d'évaluations
 - Certificatives, contexte universitaire
- Assister l'enseignant dans l'élaboration de ses sujets :
 - Différents sujets d'évaluations à produire pour différentes cohortes d'apprenants sur un même enseignement
 - Différentes filières ou encore d'une année sur l'autre.
 - Différents sujets d'évaluations à produire pour une même cohorte d'apprenants et pour un même enseignement
 - Ce cas de figure s'illustre par exemple dans les formations relevant du contrôle continu ou encore lors d'examens multisessions
 - Différents sujets d'évaluation à produire pour une unique session d'examen.
 - Limiter la fraude dans le cadre d'une composition unique en amphithéâtre; une composition par sous-groupes en horaires décalés comme par exemple dans le cas de certaines épreuves du Certificat Informatique et Internet (C2I niveau 1)



Contexte et objectifs

- Sujets différenciés
 - Disparités dans les niveaux de difficulté
 - Même si l'on considère un thème d'évaluation dans sa globalité, comment sommes-nous en mesure de garantir que chaque individu sera évalué avec un niveau d'exigence sensiblement équivalent ?
 - D'une session à l'autre ou encore d'une cohorte à l'autre.



Genèse de la recherche

- Notre première approche : Évaluations de type QCM
 - Cas maximal de sujets individuels
- **Prototype** YMCQ
 - Outil de différentiation des sujets d'examens
 - Génération automatique basée sur une approche aléatoire
 - Permet une composition sur papier
 - Automatise la génération des sujets & la correction des copies



Genèse de la recherche

- Principe de YMCQ
 - Base de questions sources
 - Énoncé + liste de réponses justes et de distracteurs pour chaque énoncé
 - Tirage au sort aléatoire des énoncés + tirage au sort des choix de réponses parmi les réponses possibles
 - Seul moyen de garantir l'équité, avoir dans les sujets générés le même nombre de questions que dans la base source → Ne faire qu'un « mélange »
 - La distance entre deux sujets générés peut être minimale



Genèse de la recherche

- Principe de YMCQ
 - Base de questions sources
 - Énoncé + liste de réponses justes et de distracteurs pour chaque énoncé
 - Tirage au sort aléatoire des énoncés + tirage au sort des choix de réponses parmi les réponses possibles

- Seul
- génér
- source
- La di

L'objectif de notre travail :
Garantir un certain niveau de différenciation
ET prendre en compte l'équité :
Le projet DIFAIRT
(Differentiated FAIRs Tests)

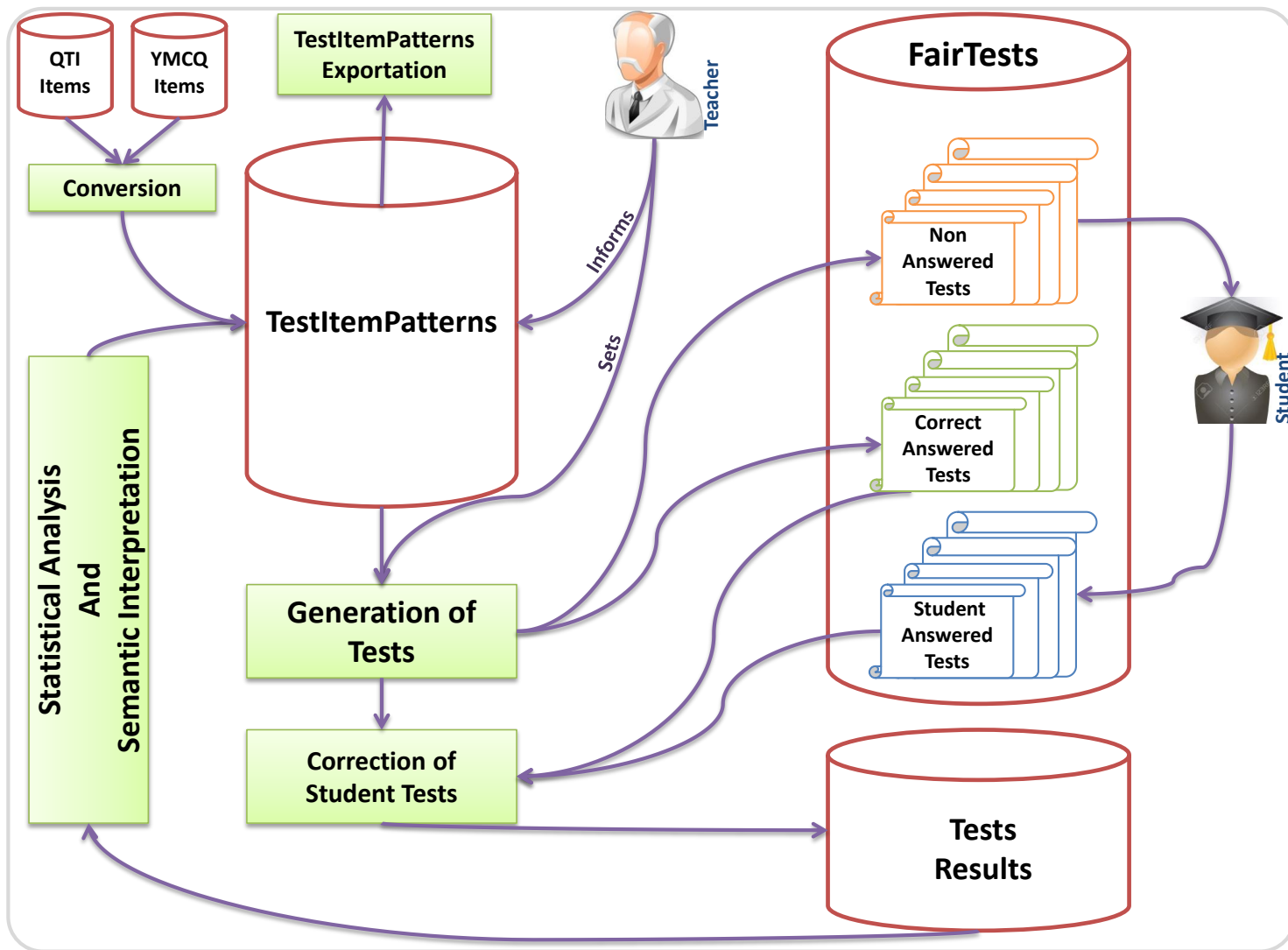


L'approche DIFAIRT

- Un test se compose d'une série d'Items (ItemsTests)
- Un ItemTest est construit à partir d'un TestItemPattern
 - L'énoncé de la question
 - Toutes les réponses possibles envisageables
 - Des données descriptives sur l'Item en général. Ex : le thème auquel se réfère la question
 - Des données descriptives sur chaque réponse possible pour l'énoncé ex: comme le degré de « plausibilité » d'un distracteur



L'approche DIFAIRT





Le générateur : DIFAIRT-G

- Algorithme de génération
 - Différentiation → Garantir une distance minimum entre chaque paire de sujets générés
 - Calcul d'une distance sur la structure du sujet
 - Maximise la disparité des énoncés de questions, et en cas de questions identiques, maximise la disparité des réponses associées
 - Equité → Minimiser l'écart type entre le niveau de difficulté de chaque paire de sujets générés
 - Mesure du niveau de difficulté d'un sujet
 - Rejet ou acceptation d'un sujet généré en fonction de ces deux dimensions



Perspectives

- Actuellement : finaliser l'implémentation du calcul de la distance entre deux sujets
 - Expérimenter pour fixer les seuils de performance
- A plus long terme : définir la mesure du niveau de difficulté d'un Test → Mesure du niveau de difficulté d'un ItemTest ?
 - A partir de données descriptives de chaque ItemTestsPattern
 - Saisies par l'enseignant puis Enrichies automatiquement à l'issue des résultats d'évaluation des étudiants
 - En fonction du contexte de l'évaluation



Perspectives

- Expérimentations envisagées de DIFAIRT-G
 - Epreuves de certifications C2i Niveau 1 à l'Université de Picardie Jules Verne
 - Formations de l'Ecole Supérieure d'Infotronique d'Haïti
- Prendre en compte la couverture thématique au moment de la génération
 - Couplage avec une ontologie du domaine d'apprentissage
- Elargir à d'autres types d'items que les QCM
- Intégrer la partie capitalisation des résultats d'évaluation des étudiants