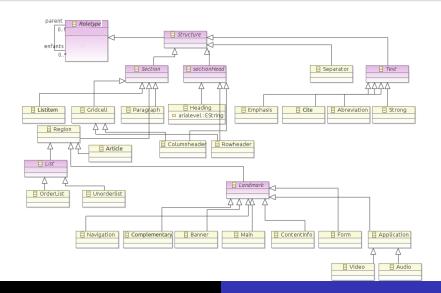
Présentation Berger Levrault

21 mai 2014

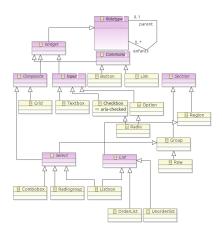
Méta-modèle éléments structuration Méta-modèle éléments d'interaction Méta-modèle éléments de mise en form nstance Méta-modèle

- Méta-modèle
 - Méta-modèle éléments structuration
 - Méta-modèle éléments d'interaction
 - Méta-modèle éléments de mise en forme
 - Instance Méta-modèle
- 2 Extraction structure
 - Démarche générale
 - Heuristiques
 - Résultats
 - Difficultés

Méta-modèle éléments structuration

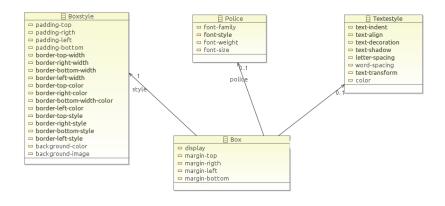


Méta-modèle éléments d'interaction



Méta-modèle éléments structuration Méta-modèle éléments d'interaction Méta-modèle éléments de mise en forme Instance Méta-modèle

Méta-modèle éléments de mise en forme



Méta-modèle éléments structuration Méta-modèle éléments d'interaction Méta-modèle éléments de mise en forme Instance Méta-modèle

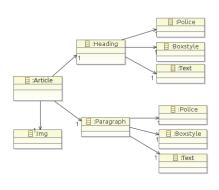
Instance méta-modèle

Loi martiale en Thaïlande : l'armée déployée, les médias censurés

MONDE 2 réactions

"Déclarer la loi martiale n'est pas un coup d'Etat", mais vise "à restaurer la paix et l'ordre public", assure l'armée.

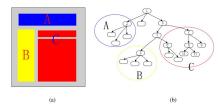




 $\label{eq:article.Heading.Police.fontSize} Article.Paragraph.Police.fontSize * 1.20) \\ Article.Heading.textStyle.color = ! "rouge" \\$

- Méta-modèle
 - Méta-modèle éléments structuration
 - Méta-modèle éléments d'interaction
 - Méta-modèle éléments de mise en forme
 - Instance Méta-modèle
- Extraction structure
 - Démarche générale
 - Heuristiques
 - Résultats
 - Difficultés

Démarche générale



Approche

- Algorithme de segmentation (descendante) basé sur des heuristiques
- Pour chaque bloc extrait (= nœuds non segmentés) est attribué un Degrés de cohérence
- Si le degrés de cohérence est faible alors on ré-applique le processus de segmentation

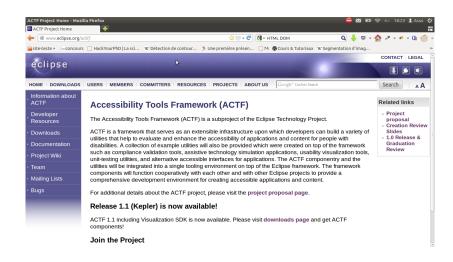
Heuristiques

- Informations de mise en forme :
 - nœuds de type breakline
 - nœuds de type inline
 - nœuds avec des propriétés de couleur
- Informations de structuration :
 - noeuds avec des tags de structuration <HR>, <P>,

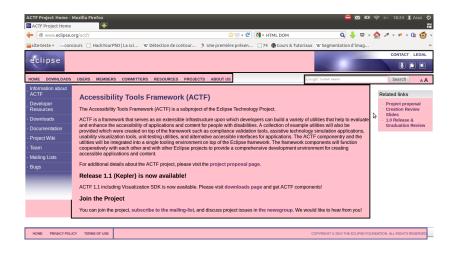
Exemples d'heuristiques :

- Segmentation des nœuds qui possèdent des enfants de type breakline
- Extraction des nœuds voisins de <HR>
- Division des nœuds dont les enfants ne partagent pas la même propriété de couleur d'arrière plan

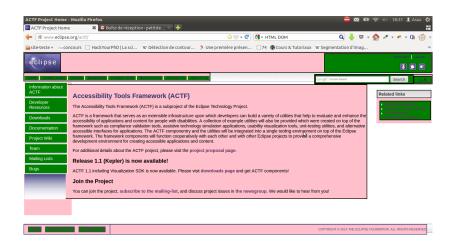
Segmentation étape 0



Segmentation étape 1



Segmentation étape 2



Segmentation trop souple



