

L3 informatique, 2120-2022, Projet de LFC

Traduction de Markdown en html

ETAPE 4 – A rendre 14 jours après votre TP3 (dates à la fin de ce fichier)

Simplification du projet

La mise en forme des textes se fera uniquement avec des étoiles (on oublie les underscores ; ils deviennent des caractères interdits).

Les étoiles « mal placées » deviennent des erreurs syntaxiques. Une étoile toute seule n'est plus considérée comme un caractère du texte et ne sera pas à traduire en html. Ceci simplifiera également la grammaire. Eventuellement, on les réintroduira plus tard en considérant qu'une étoile qui fait partie du texte doit obligatoirement être précédée d'un \ (à ne pas faire dans l'immédiat).

Pour la mise en forme des textes, on oublie également la mise en forme progressive/dégressive. Un texte est soit en italique (une étoile de chaque côté), soit en gras (deux étoiles de chaque côté), soit en gras-italique (trois étoiles de chaque côté).

On garde les états ajoutés dans le lex et on garde le stockage des morceaux de texte dans la table des symboles. Toutefois, la colonne 3 du tableau TAB est désormais inutile puisqu'on ne stocke plus les étoiles et les underscores dans la table des symboles.

Modification du programme LEX

Modifiez votre programme LEX en tenant compte de la nouvelle liste d'unités lexicales et des indications ci-dessous :

Unité lexicale	Description
Morceau de texte	Chaîne de caractères de longueur supérieure ou égale à 1, commençant par n'importe quel caractère sauf (#, *, _, \n, espace et tabulation) et pouvant contenir ensuite n'importe quel caractère sauf (#, *, _ et retour à la ligne).
Balise de titre	Chaîne de caractères positionnée en début de ligne et composée de 3 parties : une suite d'espaces (entre 0 et 3), une suite de # (entre 1 et 6), puis à nouveau une suite d'espaces (au moins 1). Ne peut être rencontrée que dans l'état INITIAL.
Fin de titre	Premier retour à la ligne ou première ligne vide rencontré(e) dans l'état TITRE.
Ligne vide	Chaîne composée d'au moins deux retours à la ligne avec éventuellement des espaces entre eux.
Début de liste	Point de liste lu quand l'analyseur lexical se trouve dans l'état INITIAL.
Item de liste	Point de liste lu quand on se trouve dans l'état ITEM.
Fin de liste	Ligne vide rencontrée quand on est dans l'état ITEM.
Etoile	Chaîne de 1 caractère : une étoile

Définition d'un point de liste : Chaîne de caractères positionnée en début de ligne et constituée d'une étoile puis d'une suite non vide d'espaces.

Définition d'une ligne vide : Chaîne composée d'au moins deux retours à la ligne avec éventuellement des espaces entre eux.

Définition d'un retour à la ligne : \n ou \r\n

L'état ITEM est l'état de l'analyseur dans lequel on passe après lecture d'un point de liste. On en ressort (on repasse dans l'état INITIAL) lorsqu'on lit une ligne vide.

L'état TITRE est l'état de l'analyseur dans lequel on passe après lecture d'une balise de titre. On en ressort (on repasse dans l'état INITIAL) lorsqu'on rencontre la fin de titre.

Les underscores ne sont plus considérés comme caractères spéciaux (plus d'unité lexicale dédiée). Ils deviennent des caractères interdits.

Les retours à la ligne doivent être ignorés s'ils ne sont pas en fin de ligne de titre (donc plus de renvoi d'unité lexicale « retour à la ligne simple ». Ceux qui se trouvent à la fin d'une ligne de titre sont associés à une nouvelle unité lexicale « Fin de titre »).

Table des symboles

Les informations à stocker dans la table des symboles sont pratiquement les mêmes que celles données lors du TP2 en ce qui concerne les morceaux de texte. Les étoiles et underscores n'ont plus besoin d'y être stockées puisque les underscores sont désormais des caractères incorrects et les étoiles ne servent plus qu'à définir une mise en forme. Du coup, la seconde colonne n'est plus utile sinon, nous aurions « texte » dans toutes les cases de cette colonne.

Fichier YACC

La grammaire décrivant la syntaxe des fichiers markdown simplifiés est la suivante :

Symboles terminaux	Symboles non terminaux
TXT (morceau de texte) BALTIT (balise de titre) FINTIT (fin de titre) LIGVID (ligne vide) DEBLIST (début de liste) ITEMLIST (item de liste) FINLIST (fin de liste) ETOILE	fichier (axiome) element titre liste suite_liste texte_formatte liste_textes
Règles fichier → element element fichier element → TXT LIGVID titre liste texte_formatte titre → BALTIT TXT FINTIT liste → DEBLIST liste_textes suite_liste suite_liste → ITEMLIST liste_textes suite_liste FINLIST texte_formatte → italique gras grasitalique italique → ETOILE TXT ETOILE gras → ETOILE ETOILE TXT ETOILE ETOILE grasitalique → ETOILE ETOILE ETOILE TXT ETOILE ETOILE ETOILE liste_textes → TXT texte_formatte TXT liste_textes texte_formatte liste_textes	

Vous devez maintenant écrire le fichier YACC qui correspond à cette grammaire et le relier au fichier LEX puis terminer le stockage des morceaux de texte dans la table des symboles.

Vous pourrez tester votre analyseur lexical et syntaxique sur le fichier Exemple2.txt fourni.

Les résultats à obtenir sont sur les pages suivantes

Affichages du lex et du yacc (en rouge, affichage du yacc)

Morceau de texte : Exemple de texte de test

Ligne vide

Balise de titre

Morceau de texte : avec un titre de niveau 1

Fin de titre

Balise de titre

Morceau de texte : et un titre de niveau 2

Fin de titre

Début de liste

Morceau de texte : une liste comportant 3 items

Item de liste

Morceau de texte : le second avec un texte non formaté

Item de liste

Morceau de texte : et le troisième avec un texte

Etoile

Morceau de texte : en italique

Etoile

Fin de liste

Morceau de texte : puis un texte

Etoile

Etoile

Morceau de texte : en gras

Etoile

Etoile

Morceau de texte : et un autre

Etoile

Etoile

Etoile

Morceau de texte : en gras italique

Etoile

Etoile

Etoile

Ligne vide

Morceau de texte : ici un caractère interdit :

Erreur lexicale : Caractère _ non autorisé

Morceau de texte : et ici un texte

Etoile

Morceau de texte : syntaxiquement incorrect

Ligne vide

erreur syntaxique

La table des symboles

TAB

0	0	24	Normal
1	24	25	Titre
2	49	23	Titre
3	72	28	Item
4	100	37	Item
5	137	31	Item
6	168	12	Item
7	180	14	Normal
8	194	8	Normal
9	202	12	Normal
10	214	17	Normal
11	231	29	Normal
12	260	16	Normal
13	276	25	Normal

CH

Exemple de texte de test avec un titre de niveau 1 et un titre de niveau 2 une liste comportant 3 items le second avec un texte non formaté et le troisième avec un texte en italique puis un texte en gras et un autre en gras italique ici un caractère interdit : et ici un texte syntaxiquement incorrect

Travail à rendre au plus tard 14 jours après votre TP2, soit

- Groupe TP1 : 25 mars 23h55
- Groupe TP2 : 24 mars 23h55
- Groupe TP3 : 28 mars 23h55
- Groupe TP4 : 24 mars 23h55
- Groupe TP5 : 24 mars 23h55
- Groupe TP6 : 23 mars 23h55

Fichiers à rendre sur plubel en utilisant le lien « Dépôt étape 3 projet ». Pensez à mettre votre numéro de groupe dans l'un des fichiers. Ajoutez un makefile ou un fichier texte dans lequel vous donnerez les commandes de compilation. Regroupez tous vos fichiers (lex, yacc, etc.) dans une archive zip et rendez un fichier par groupe de projet.