

Rapport technique
Défi Fouille de Textes

Stefana Bolohan - p1095957
Pauline Tremblay – p1096019

© 2014 par Stefana Bolohan et Pauline Tremblay. Ce travail a été réalisé à l'EBSI, Université de Montréal, dans le cadre du cours INU3051 – Information et sites Web donné au trimestre d'automne 2014 par Dominic Forest (remis le 28 novembre 2014).

Historique des révisions

Version	Date	Description	Auteurs
1	25/10/2014	- Début de rédaction du but, de la portée et des acteurs	Flora Fislewicz
2	30/10/2014	- Rédaction et finalisation du but	Pauline Tremblay
3	03/11/2014	- Rédaction et finalisation de la portée - Intégration de la structure du site - Début de rédaction de la trame des pages	Stefana Bolohan Pauline Tremblay
4	04/11/2014	- Intégration du planning - Rédaction et finalisation des acteurs	Stefana Bolohan Pauline Tremblay
5	05/11/2014	- Intégration du dessin de la trame - Rédaction et finalisation de la trame des pages	Stefana Bolohan Pauline Tremblay
6	24/11/2014	- Début de la rédaction de la charte graphique.	Pauline Tremblay
7	26/11/2014	- Rédaction de la charte graphique	Stefana Bolohan
8	27/11/2014	- Fin de la rédaction de la charte graphique - Insertion de la maquette du site	Stefana Bolohan Pauline Tremblay

Note : Ce document est un outil de travail. Il est inspiré de plusieurs ouvrages, dont les suivants :

- Bordage, S. 2006. Conduite de projet Web, 3e éd. Paris : Eyrolles.
- Chu, N. 2006. Réussir un projet de site Web, 4e éd. Paris : Eyrolles
- Goto, K. et Cotler, E. 2005. Web redesign. Workflow that works, 2nd ed. Indianapolis : New Riders.
(cet ouvrage est aussi disponible en français sous le titre Redesign Web 2.0. Conduite de projet publié chez Eyrolles)

Table des matières

Historique des révisions	1
Table des matières	3
1. Présentation générale du projet.....	4
1.1. But	4
1.2. Portée	4
1.3. Définitions, acronymes et abréviations	5
1.3.1. Définitions	5
1.3.2. Acronymes et abréviations	5
1.4. Références	5
1.5. Sommaire	5
2. Acteurs (fonction / description)	5
3. Planification	6
4. Structure du site et des pages Web.....	7
4.1. Plan du site.....	7
4.2. Trame des pages.....	7
5. Spécifications techniques	13
5.1. Règles de nommage et de codage.....	13
6. Maquettes et modèles graphiques.....	14
7. Charte graphique.....	14
7.1. Dimensions.....	14
7.2. Logo	15
7.3. Polices et jeux de caractères.....	15
7.4. Jeux de couleurs	16
7.5. Autres éléments graphiques (incluant les traitements).....	16
8. Stratégies de référencement utilisées (ou à recommander)	17

1. Présentation générale du projet

1.1. But

À la demande du DÉfi Fouille de Textes, le projet a pour but la fusion de trois sites internet du DÉfi en un site web intégré entièrement conforme aux normes HTML5 et CSS3 du W3C. Les trois sites suivant seront rassemblés: le portail général (<http://deft.limsi.fr/>), la page du DEFT2014 (<http://deft.limsi.fr/2014/>), qui est l'année en cours de l'atelier, et la page du DEFT2013 (<http://deft.limsi.fr/2013/>). Le client attend également des concepteurs la création d'une nouvelle identité visuelle.

Conformément aux attentes du client, le projet devra répondre à des objectifs à la fois fonctionnels et visuels, qui sont les suivants:

1. Mettre en valeur du contenu du site et des activités du DÉfi Fouille de Textes, en uniformisant, en réorganisant et, au besoin, en modifiant certains contenus.
2. Permettre à l'utilisateur du site d'accéder le plus facilement possible à toutes les ressources en intégrant plusieurs éléments d'architecture de l'information (système de navigation global, fil d'Ariane, système de navigation local, etc.).
3. Créer un site web sobre, qui reflète à la fois le sérieux et l'originalité de l'organisation.
4. Créer une identité visuelle unique et uniforme qui rendra le site attractif.
5. Améliorer le référencement du site.

1.2. Portée

Les contenus du site sont spécialisés, et concernent également les informations relatives aux ateliers. Les personnes le visitant chercheront donc probablement à connaître le lieu et la thématique du prochain atelier, les modalités d'inscription, le programme de la campagne en cours, ou encore les ressources des campagnes antérieures. À partir de ces constatations, plusieurs catégories d'utilisateurs du site du DÉfi Fouille de Textes peuvent être identifiées:

1. Les personnes participants au DÉfi Fouille de Textes.
2. Les chercheurs ou professionnels en fouille de textes (regroupant donc les domaines de l'informatique, de la linguistique, ou encore des statistiques).
3. Les étudiants de ces domaines s'intéressant à la fouille de textes.
4. Les membres des institutions partenaires (TALN, etc).
5. Le public général (amateurs) intéressé par la fouille de documents.

Une analyse de la concurrence peut s'avérer utile dans le cas d'une refonte de site web. Cependant, après avoir effectué une recherche sur d'autres ateliers ou conférences en fouille de textes, il n'existe pas d'organisation du même type dans le domaine francophone. Cependant, l'atelier du DÉfi Fouille de Textes fait souvent partie de la conférence annuelle francophone Traitement Automatique du Langage Naturel (TALN) dont le site est plus moderne, plus intelligible et plus accessible que ceux du DEFT actuellement. Il est donc nécessaire d'apporter des changements à ce dernier.

Dans une logique de mise en valeur des ressources du site, le visiteur devra donc avoir accès à chacune des sections suivantes:

- La campagne en cours (DEFT2014).
- Les campagnes antérieures (ici, seulement le DEFT2013).

- Les actes de chaque année (2005 - 2014).
- Les corpus de chaque année (2005 - 2014).
- Les partenaires au fil des années.

1.3. Définitions, acronymes et abréviations

1.3.1. Définitions

1.3.2. Acronymes et abréviations

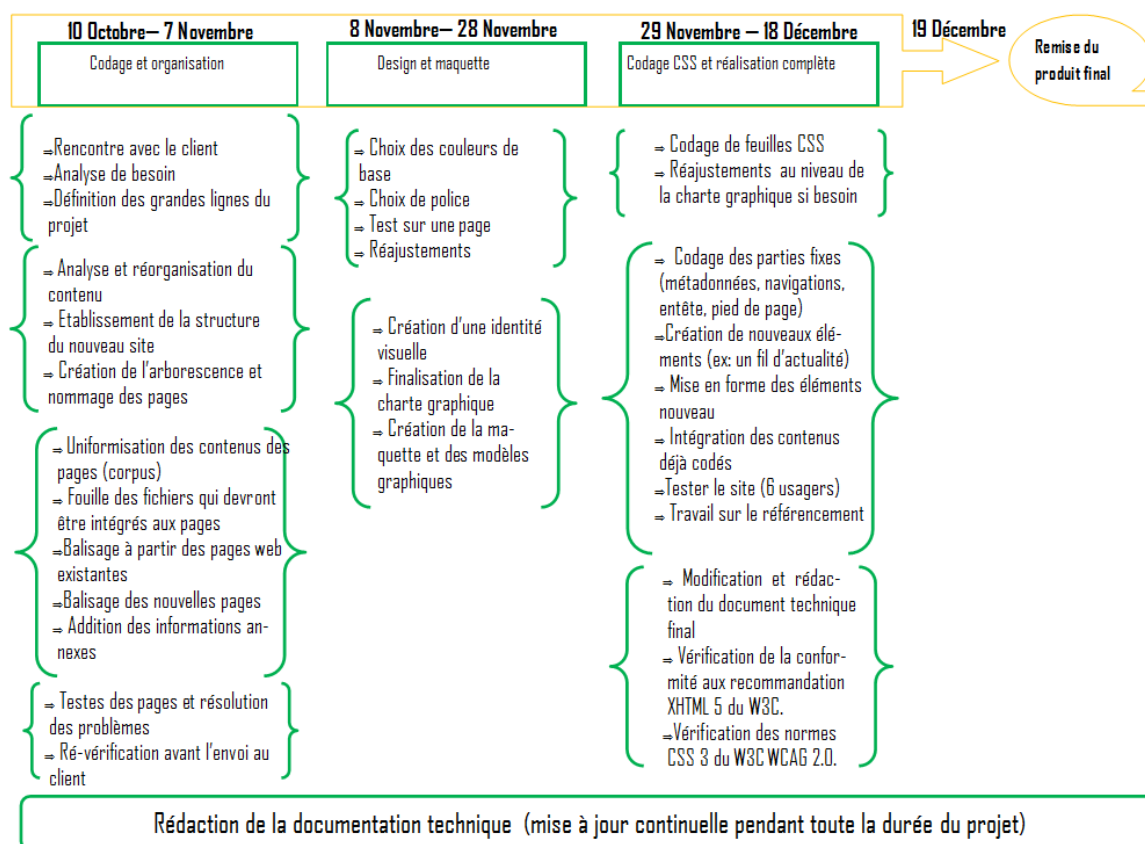
1.4. Références

1.5. Sommaire

2. Acteurs (fonction / description)

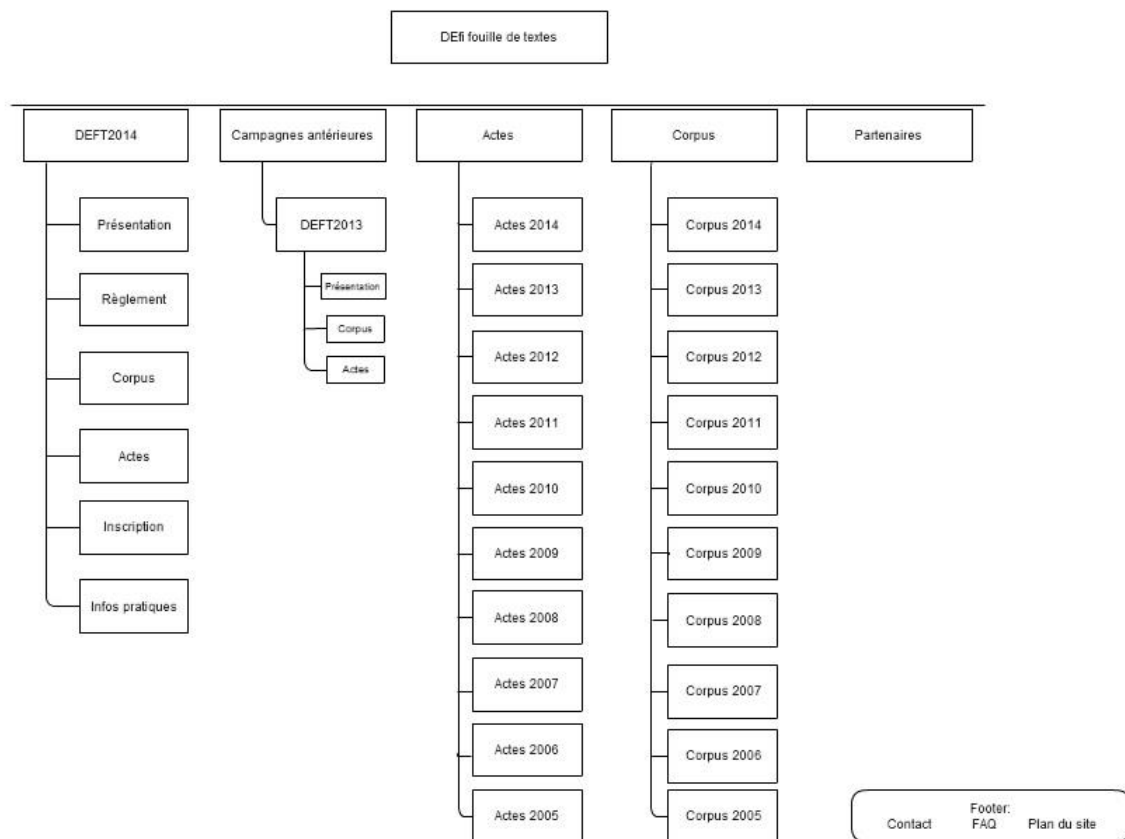
Fonctions	Responsable(s)
Client <ul style="list-style-type: none"> - Définir le site web et les besoins - Validation du projet 	DEFT
Agent de liaison	Dominic Forest
Chef de projet et ingénieur réseau <ul style="list-style-type: none"> - Établir les réunions et les tâches - Mettre en place une infrastructure permettant le travail à distance ainsi que le partage des documents - Faciliter la communication au sein du groupe - Rassembler et mettre à jour des versions de documents - nommage de page et réorganisation du contenu - Créer l'arborescence - Mettre les documents sur le serveur et communiquer avec le commanditaire 	Stefana Bolohan
Chargé d'étude et architecte Web <ul style="list-style-type: none"> - Définir les objectifs - Tester, identifier et traiter les dysfonctionnements éventuels - Organiser les contenus des pages web et leur uniformité - Vérifier les pages pour leur conformité avec les normes en vigueur - Étudier les sites qui peuvent porter sur la même thématique et identifier les éléments essentiels pour le projet - Définir la maquette du site web 	Pauline Tremblay
Concepteur/Infographiste/Développeur Web <ul style="list-style-type: none"> - Réfléchir sur les informations à inclure dans le nouveau site web - Procéder au codage du contenu en XHTML strict - Valider les codes sources à l'aide du validateur du W3C - Tester le fonctionnement du site sur différents navigateurs - Déterminer la charte graphique du site web - Créer des feuilles de style en CSS 3 - Rédiger le rapport technique 	Stefana Bolohan Pauline Tremblay

3. Planification



4. Structure du site et des pages Web

4.1. Plan du site



4.2. Trame des pages

4.2.1. Éléments récurrents

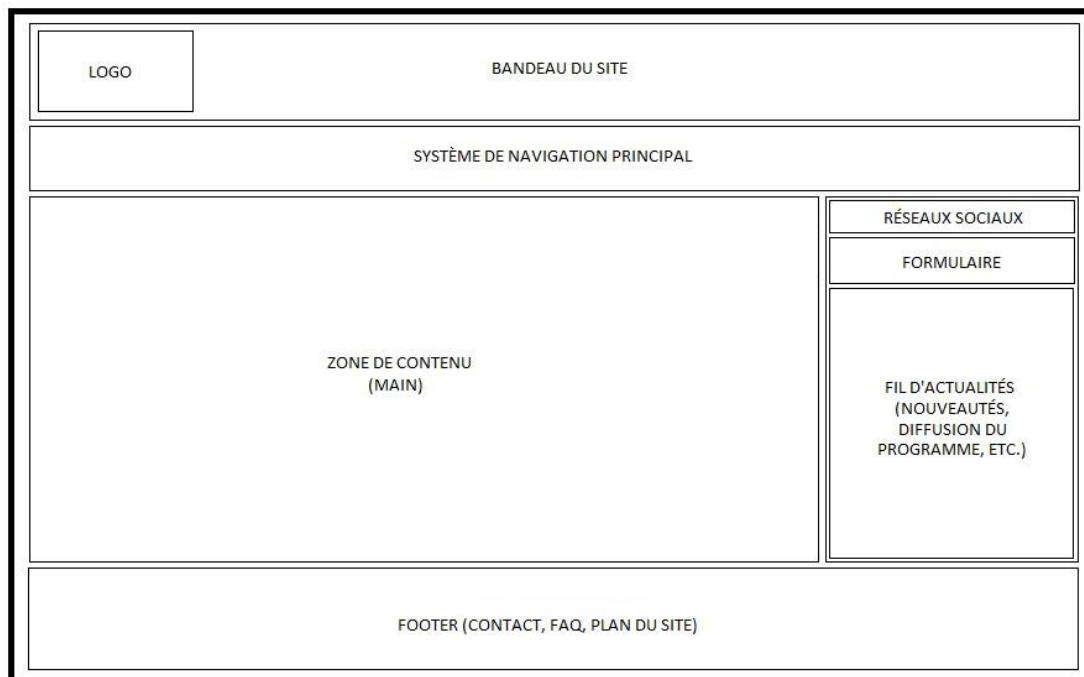
Conformément à la volonté du client de procéder à une uniformisation des pages et à une réorganisation dans le but de rendre le site intelligible, les éléments suivants seront récurrents :

- Éléments d'identification dans le Header: logo et nom de l'organisme.
- Éléments de navigation:
 - Système de navigation global,
 - Fil d'Ariane,
 - Contact, FAQ et Plan du site dans le Footer.
- Un Aside qui comportera les éléments suivants:
- Boutons de partage sur les réseaux sociaux,
- Formulaire de renseignement de l'adresse mail, pour que l'utilisateur puisse être tenu au courant des actualités du DEFT sans avoir à visiter le site (la présence de cet outil peut pousser le

détenteur du site à créer plus d'informations et plus de contenus, et donc à rendre son site plus vivant).

- Un fil d'actualités.

Voici un schéma de la trame de pages présentant les éléments récurrents :



4.2.2. Page d'accueil

Titre de la page	Nom du fichier
Accueil	homepage.html

La page d'accueil comporte les informations essentielles sur le DEFT. Une reformulation de la présentation du DÉfi (description et objectifs) a été faite. La redirection de l'utilisateur vers la notice de l'ouvrage sur le DÉfi a été conservée et reformulée. La liste des différentes éditions du DEFT, avec la description rapide de leur thématique a été conservée. Les autres éléments initialement présents sur la page d'accueil de l'ancien site ont été soit supprimés, soit déplacés pour plus d'intelligibilité.

4.2.3. DEFT2014

Titre de la page	Nom du fichier
Présentation	cec_desc.html
Règlement	cec_regle.html
Corpus	crp_2014
Actes	actes_2014
Inscription	inscription.html

Informations pratiques	cec_infos.html
------------------------	----------------

a. Présentation

La présentation du DEFT2014 comporte trois éléments:

- La description générale de la thématique de la campagne en cours.
- La description des tâches de la campagne en cours.
- La liste du comité d'organisation et de programme.

Plutôt que de séparer les différents comités, qui comportaient les mêmes personnes, il a été préférable d'indiquer, dans une courte phrase, que le comité d'organisation s'occupe également de la programmation.

b. Règlement

Le choix a été fait de séparer le règlement de la présentation du DEFT2014. En effet, le règlement s'adresse plus particulièrement aux personnes participant au DÉfi puisqu'il décrit les différents impératifs des participants (impératifs liés aux tâches et aux corpus).

c. Corpus 2014

Toutes les éditions DEFT contiennent des corpus sur lesquels les participants ont travaillé. Les corpus de 2014 sont un particulier importants car il s'agit de l'édition en cours. Les éléments comme le but, une description de type de ressources utilisés ainsi que les dates essentielles pour l'accès aux corpus sont présents en début de page. Les tâches sont ici développées et associées aux exemples. Une particularité qui n'apparaît que sur la page de corpus 2014 est la partie qui présente les formats de sortie et évaluation associés aux tâches. Comme sur toutes les pages de corpus la fin de la page donne la place aux fichiers de téléchargement contenant les corpus. Cette page contient beaucoup d'information mais nous considérons important de donner à l'utilisateur une documentation complète, surtout pour l'année en cours.

d. Actes 2014

Les informations concernant les actes 2014 sont inclus dans la description des corpus ci-dessous.

e. Inscription

Étant donné que la période d'inscription est arrivée à son terme, le choix a été fait d'indiquer, comme sur le site d'origine, que les inscriptions étaient closes. Il aurait été contre-productif de créer un formulaire d'inscription, en indiquant au-dessus du formulaire ou à la suite de la validation par l'utilisateur que les inscriptions n'étaient plus d'actualité.

f. Informations pratiques

Les informations pratiques permettent au visiteur de connaître le lieu de la conférence grâce à l'indication de l'adresse (utile pour les personnes se rendant à la conférence en voiture), mais aussi grâce à une carte générée sur Google.

Il est également d'usage d'indiquer les informations nécessaires quant aux transports en commun: gare à proximité, ligne de métro et lignes de bus desservant le lieu de la conférence. Comme complément d'information, l'utilisateur trouvera aussi un lien de redirection vers le site des transports publics de Marseille.

4.2.4. Campagnes antérieures

Titre de la page	Nom du fichier
Présentation	ca_desc.html
Corpus	crp_2013.html
Actes	actes_2013.html

a. Présentation du DEFT2013.

La présentation du DEFT2013 comporte une description ainsi que la liste des personnes du comité d'organisation et du comité de programme. Comme pour les comités de 2014, les noms de chaque comité étaient les mêmes, à l'exception de deux personnes. Plutôt que de créer deux listes, qui auraient alourdi la page, le parti a été pris d'indiquer, par une courte phrase, les deux personnes supplémentaires au sein du comité de programme.

b. Corpus 2013

La page de corpus 2013 est plus allégée que les autres page de corpus. Elle contient tous les éléments communs aux pages de ce type (cf. structure commune des corpus) à l'exception des corpus utilisés et des exemples. L'impossibilité de fournir des informations sur ces deux volets fait que l'accent est mis surtout sur les tâches. Les tâches servent aussi comme présentation des ressources utilisées. En bas de page l'indisponibilité de corpus est annoncée.

c. Actes 2013

Les informations concernant les actes 2013 sont inclus dans la description des corpus ci-dessous.

4.2.5. Actes

Titre de la page	Nom du fichier
Actes 2014	actes_2014.html
Actes 2013	actes_2013.html
Actes 2012	actes_2012.html
Actes 2011	actes_2011.html
Actes 2010	actes_2010.html
Actes 2009	actes_2009.html
Actes 2008	actes_2008.html
Actes 2007	actes_2007.html
Actes 2006	actes_2006.html
Actes 2005	actes_2005.html

Toutes les pages des actes sont construites sur le même modèle:

- Une version de l'intégral des actes est proposée.
- Les articles de présentation de l'édition et ses résultats (s'il y a lieu).
- Une liste des autre articles, présentant également le ou les auteur(s), l'affiliation et quelques mots-clés.

L'ancien site présentait les actes par articles, auteurs, affiliations et mots-clés dans des tableaux, ce qui alourdissait les pages, pour finalement rediriger l'utilisateur vers le même contenu. Il a été préférable de créer une liste, qui sera ensuite stylisée. Le titre de l'article, qui renvoie vers le fichier pdf associé, apparaît sur une première ligne, puis viennent le ou les auteur(s) et l'affiliation sur une seconde ligne, et enfin, les mots-clés sur une troisième. Alors que l'information était auparavant dispersée sur quatre pages, le visiteur peut désormais avoir accès à l'ensemble sur une seule, ce qui répond adéquatement à la demande du client, qui souhaitait que le nouveau site mette en valeur le contenu, et facilite la recherche du visiteur.

4.2.6. Corpus

Titre de la page	Nom du fichier
Corpus 2014	crp_2014.html
Corpus 2013	crp_2013.html
Corpus 2012	crp_2012.html
Corpus 2011	crp_2011.html
Corpus 2010	crp_2010.html
Corpus 2009	crp_2009.html
Corpus 2008	crp_2008.html
Corpus 2007	crp_2007.html
Corpus 2006	crp_2006.html
Corpus 2005	crp_2005.html

Les corpus regroupent des informations sur les différents matériaux utilisés dans les ateliers DEFT au fil des années. En effet, avoir un panorama de tous les corpus depuis la première édition de l'atelier en 2005 permet à l'utilisateur:

- D'observer les différentes thématiques et les traitements appliqués à celles-ci.
- De se rendre compte de l'évolution de cet atelier et de son existence annuelle.
- De comparer les corpus et les tâches à faire d'une édition à l'autre.
- D'observer le caractère de l'atelier et sa notoriété.

Enfin, la présence des corpus peut s'avérer un élément incitateur de participation au DEFT ainsi qu'un matériel précieux pour accompagner les actes et donner la possibilité de refaire les expériences.

Les pages de corpus ont une structure commune qui se retrouvent dans toutes les pages et des éléments spécifiques à chaque corpus.

La structure commune à tous les pages contient:

- Le but de chaque édition du DEFT.
- Une description des thématiques des corpus ainsi que les types de ressources utilisés.
- Les types de corpus associés aux exemples pratiques.
- Les tâches adaptées aux corpus.

Les éléments spécifiques des pages peuvent être:

- Des données supplémentaires (ex: index de données de test pour le corpus de 2006, les classes correctes pour le corpus de 2007, données post-workshop pour le corpus de 2011 etc).
- Des spécifications sur les formats des corpus.
- La disponibilité des fichiers à télécharger (pour 2005, 2006, 2007 et 2014)/ leur indisponibilité (pour les autres années).

Car les corpus font partie des archives, le contenu a été révisé et des changements ont été apportés dans la formulation pour que les verbes soient au passé et que des informations qui ne sont plus valables (ex: date) n'apparaissent pas. Tous ces changements ont été faits à partir des informations recueillies depuis les anciennes pages de corpus.

4.2.7. Partenaires

Titre de la page	Nom du fichier
Partenaires	partenaires.html

La page des partenaires est une présentation exhaustive de logo des partenaires avec qui l'atelier de DEFT a été associé au fil des années. Le but de cette page est de montrer pratiquement les relations institutionnelles que cette campagne d'évaluation en fouille de textes a construites. Cet événement annuel regroupe 10 éditions tenues dans des endroits différents (France, Québec et Suisse). Chaque année plusieurs partenaires soutiennent le DEFT, et il se trouve que certains accompagnent ce défi durant plusieurs éditions. Étant donné ce caractère répétitif, une présentation chronologique des partenaires ferait apparaître le même logo plusieurs fois. Pour résoudre cet inconvénient une analyse a été réalisée pour décider d'un ordre de priorité. L'observation des occurrences de présence de partenaires a fait ressortir un ordre de priorité (ex: TALN et LIMSI sont les partenaires le plus présents tandis que le LIRMM n'est partenaire que pour une édition). Une fois établi cet ordre, un deuxième ciblage a été effectué. Ceux qui se trouvaient sur la même position (selon le critère de priorité) ont été présentés par ordre chronologique de plus récent au plus ancien. Le commanditaire désire que les partenaires soient visibles et de ce fait des modifications visuelles seront envisagées par la suite.

4.2.8. Pages du footer

Titre de la page	Nom du fichier
Contact	contact.html
Foire aux questions	faq.html
Plan du site	plan_site.html

a. Contact

La page de contact sera accessible via le footer. Plutôt que de simplement indiquer l'adresse mail du DEFT (il aurait été peu pertinent de créer une page à cette fin), la solution de créer un formulaire d'envoi direct vers cette adresse a été privilégiée. L'utilisateur peut donc contacter le détenteur du site en indiquant ses nom et prénom, son adresse mail, l'objet de son message, et son message. L'utilisation d'un formulaire de contact rend le site plus dynamique, et facilite également la navigation de l'utilisateur au sein des différentes pages web qu'il visite. En effet, il peut ainsi rester sur la page, au lieu d'ouvrir sa boîte mail.

b. Foire aux questions

La FAQ reprend la FAQ originelle du site. Cependant, afin d'alléger la page, les termes récurrents "Question" et "Réponse" ont été supprimés. Un autre système de mise en valeur des éléments sera utilisé lors du codage CSS (caractères gras ou italiques, couleurs, etc.).

Aucun formulaire de question supplémentaire n'a été fait, mais l'utilisateur est redirigé vers la page Contact s'il souhaite poser une autre question.

c. Plan du site

Le plan du site permettra aux visiteurs de s'orienter. L'utilisation d'un plan de site répond à la demande du client de faciliter la navigation de ses visiteurs.

5. Spécifications techniques

5.1. Règles de nommage et de codage

6. Maquettes et modèles graphiques



7. Charte graphique

7.1. Dimensions

Une très grande majorité des utilisateurs possèdent une résolution d'écran minimale de 1024x768 ou plus. Dans le but de s'adapter à un maximum de visiteurs, la largeur du container sera d'une taille fixe de 960px.

Composant de la page	Dimensions
Conteneur général	Largeur – 960px Hauteur – inherit
Header	Largeur – 960px Hauteur – 200px
Système de navigation globale	Largeur – 960px Hauteur – 45px
Fil d'Ariane	Hauteur – 40px
Main	Largeur – 620px - Padding 10px à gauche. - Margin 10px à droite. Hauteur – inherit
Aside	Largeur – 310px - Margin 10px à gauche. Hauteur – inherit
Footer	Largeur – 960px Hauteur – 100px

7.2. Logo

Un logo plus moderne a été créé à l'occasion de la mise en place du nouveau site web de l'organisme. Le logo sera cliquable et agira comme un renvoi vers la page d'accueil du site.



- **Dimensions:** 180 px
- **Police:** le logo sera le seul élément du site, avec le texte du header, possédant une police d'écriture différente de celle du reste du contenu. La police choisie est Garamond.

7.3. Polices et jeux de caractères

Afin d'assurer la lisibilité et l'accessibilité du site, plusieurs choix ont été faits quant à la question des polices:

- Une police sans serif sera utilisée sur l'ensemble du site.
- La police choisie est la police Open Sans. Dans le code CSS, il sera précisé que les polices de remplacements seront en premier lieu Arial, en deuxième lieu Verdana.
- La taille minimum de la police sera de 12px.
- Les liens (à l'exception du système de navigation globale, du système de navigation locale) sont soulignés.

Composant de la page	Police	Taille	Particularités
Body	Open Sans, sans serif	12px	-
h1	Open Sans, sans serif	18px	Gras
h2	Open Sans, sans serif	16px	Gras
h3	Open Sans, sans serif	14px	Gras
Header (texte)	Garamond	40px	-
Liste dans le main	Open Sans, sans serif	12px	-
Liste dans l'aside	Open Sans, sans serif	12px	-
Fil d'Ariane	Open Sans, sans serif	12px	-
Navigation	Open Sans, sans serif	14px	Gras
Footer	Open Sans, sans serif	14px	-

7.4. Jeux de couleurs



Tous les assemblages de couleur ont été sujets à une évaluation du contraste à l'aide du site http://snook.ca/technical/colour_contrast/colour.html, afin de respecter les règles d'ergonomie. Tous les agencements sont donc conformes à ces règles. Il est à noter que le bleu a été utilisé aussi en transparence. Derrière le conteneur principal, une image de fond sera insérée.

Composant de la page	Arrière-plan	Texte	Liens
Container général	# FFFFFFFF	-	-
Header	# FFFFFFFF	#A01B2E	-
Navigation globale	#252A47 #A01B2E lorsque le curseur est sur la case	#FFFFFF	-
Main	# FFFFFFFF	p : #000000 h1/2/3 : #252A47	Visité : #A01B2E Non-visité : #252A47
Aside	# FFFFFFFF	#000000 pour p et li # FFFFFFFF pour h1	Visité : #A01B2E Non-visité : #252A47

	#252A47 en transparence pour le titre du fil d'actualité		
Footer	#252A47	#D9D9D9	#D9D9D9

7.5. Autres éléments graphiques (incluant les traitements).

Réseaux sociaux.

Les boutons de réseaux sociaux sont placés au-dessus du fil d'actualités. Il est important que les usagers puissent partager l'information relative au DEFT. Pour permettre cela, des boutons Facebook, Twitter seront intégrés à la page et un fil RSS sera créé.

Partenaires.

La majorité des images du site sont des logos des partenaires du DEFT au fil des années. Ces images seront redimensionnées et intégrées à un diaporama dynamique, présent sur la page d'accueil du site. Il se situera donc dans le main.

- Dimensions : 550px de largeur et 350px de hauteur au centre en bas du main de la page d'accueil.
- Couleurs :
 - Le cadre du diaporama est en bleu (#252A47), mais en transparence, tout comme le cadre d'affiche du nom des partenaires et les flèches de direction.
 - Le titre est en bleu (#252A47).
 - Le nom des partenaires apparaît en blanc (FFFFFF).

8. Stratégies de référencement utilisées (ou à recommander)