OZALID

Mise en place de la 1ère expérimentation

Mars / mai 2013 : 1ère expérimentation de la plateforme OZALID

Objectifs et enjeux :

- Tester le fonctionnement global de la plateforme OZALID
- Tester les fonctionnalités de correction d'OCR
- Avoir des retours sur les IHM

Principes retenus :

- Accès en mode authentifié et Online
- 2 supports envisagés : sur PC et sur tablette (Android 10 pouces)
- Sur un panel maîtrisé (Étudiants de Paris 8 + en interne à la BnF ou chez Orange)

Calendrier :

- Mars 2013 à la BU de Paris : Observation des comportements en situation de correction
- Mai 2013 au Labo de la BnF : Observation des activités en réseau

2 phases d'expérimentation / 2 évènements

1/ Observation des comportements en situation de correction

Objectifs :

- Tester les fonctionnalités de correction d'OCR
- Avoir des retours sur les IHM du module de correction

Format

- Evénement à BU Paris 8
  - Lancer l'expérimentation
  - Faire tester le support tablette
  - Recruter des testeurs pour l'expérimentation
- Expérimentation scénarisée (sur 1 semaine)
  - 1 heure d'entretien filmé pour l'observation des comportement en situation de correction

Panel

- Panel restreint (15 pers. max)
- Profil : étudiant de Paris 8 (Master Archives)

2/ Observation des activités en réseau

Objectifs :

- Tester le fonctionnement global de la plateforme
- Suivi des activité en réseau (organisation et partage de la correction, suivi des affinités, mesure de l'implication)

Format

- Evénement au Labo de la BnF
  - Présentation du projet OZALID (Atelier Labo)
  - Workshop de correction sur tablette
- Expérimentation en réseau
  - Réalisée simultanément dans plusieurs lieux sur un créneau de 1 à 2h.
  - Une salle dédiée avec des postes PC à la BnF
  - Séance animée par un ergonome de Paris 8

Panel

- Panel élargi (30 pers min.)
- Profil : non encore défini

Bibliothèque Universitaire de Paris 8







- Possibilité d'accueillir un atelier (salle des enseignants-chercheurs) ou une promenade numérique à la rencontre des étudiants
- Présentation du projet en parallèle de l'expérimentation
- Outils de communication :
  - ✓ Affichage
  - ✓ Mailing liste
  - ✓ Réseaux sociaux de l'université et de la BU



Le Labo de la BnF





Espace expérimental de 80 m2 dédié aux nouvelles technologies de l'information et de la communication et aux innovations techniques liées à la lecture et à l'écriture

- Lieu d’exposition
- Organisation d’Atelier
- Borne Wifi

Blog



Fil twitter



Préparation de l'expérimentation

Ressources humaines

- Un ergonome de Paris 8 (responsable de la mise en place de l'expérimentation) à plein temps pendant 2 mois minimum pour :
  - Etablir la grille d'observation et le questionnaire de l'entretien
  - Réaliser les entretiens d'auto confrontation
  - Animer l'expérimentation en réseau
  - Analyser les données
- Recrutement d'un stagiaire par Urbilog pour coordonner l'organisation des événements

Matériel

- 5 à 8 tablettes pour les opérations événementielles (Orange et la BnF)
- 3 Postes PC pour toute l'opération (Orange)
- 4 Caméras vidéo (Paris 8)
- Salle dédiée équipée de PC pour l'expérimentation en réseau (la BnF)
- Outils de communication (en cours d'élaboration)

Travaux en cours

- Elaboration d'une grille d'évaluation pour définir « a priori » le niveau de difficulté des documents
- Travail avec le Groupe Traces pour l'analyse des données suite à l'expérimentation

Définition du Corpus
Grille de critères techniques

LOT 1 : documents de complexité faible

- Structure : mono-colonne, peu de titres
- Typographie : peu d'italiques ou de gras
- Résolution. 300dpi ou 400dpi
- Colorimétrie : couleur et niveaux de gris.
- Taux de reconnaissance ABBY : moitié > 95%, moitié entre 80% et 95%.

- Monographie après 1820 (hors dictionnaires, théâtre, poésie et ouvrage scientifique)
- Provenant des programmes de numérisation de masse (postérieurs à 2007 et/ou à 2011)
- Taux de qualité d'OCR : > 95%, et entre 80% et 95%.

LOT 2 : étalon pour estimer les limites du prototype

- Structure : sans contrainte
- Typographie : sans contrainte.
- Résolution. sans contrainte.
- Colorimétrie : sans contrainte.
- Taux de reconnaissance : sans contrainte

- Panel de journaux, revues, livres anciens, ouvrages scientifiques, dictionnaires, théâtre et poésie
- Provenant de programmes différents (antérieurs et postérieurs à 2007)

Définition du Corpus
Sélections documentaires thématiques

Romans populaires : Collection "Patrie"
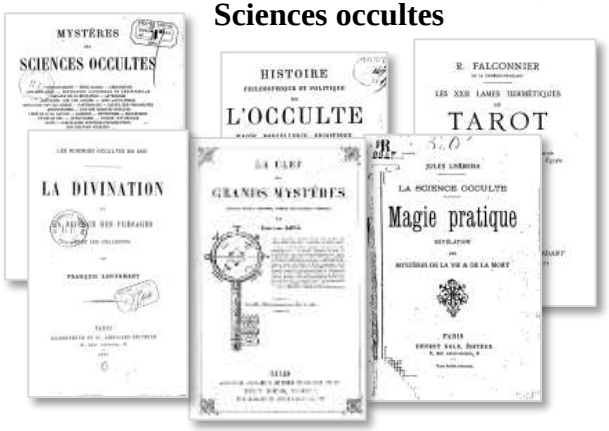


Manuels Roret : Série de manuels de vulgarisation des sciences, des arts et métiers, etc.



Romans populaires : Collection "Crimes et châtiments"



Sciences occultes



Documents les plus consultés



Grille d'évaluation des documents

	Qualité du document					Qualité de l'OCR				Nature du document							Classement	
	0,4	0,2	0,2	0,2	Note	0,3	0,5	0,2	Note	0,1	0,2	0,3	0,3	0,1	Note	Note finale		
Titre	Aspect visuel	Résolution	Couleur	Support		OCR prest.	OCR réel	Dispersion		Taille	Densité car.	Nbre total car	Nbre pages	Date d'édition			seuils fixes	écart-type
Atlas des champignons comestibles et vénéne	2	3	3	3	**2,6**	3	3	2	**2,4**	1	2	3	3	2	**2,5**	**2,50**	facile	facile
Art d'être grand-père	2	2	1	3	**2**	3	3	3	**2,4**	2	3	2	0	2	**1,6**	**2,00**	moyen	moyen
Gustave Flaubert, 1821-1880 : sa vie, ses roma	2	3	1	1	**1,8**	3	3	3	**2,4**	2	2	0	0	2	**0,8**	**1,67**	moyen	difficile
La Vie végétale, histoire des plantes à l'usage	2	2	1	3	**2**	1,5	3	3	**1,95**	3	2	0	0	2	**0,9**	**1,62**	moyen	difficile
Pel, l'horloger empoisonneur : un ancêtre de La	1	3	1	1	**1,4**	1	2	1	**1,3**	3	2	2	3	2	**2,4**	**1,70**	moyen	moyen
Mata-Hari	1	3	1	1	**1,4**	2	2	2	**1,6**	3	2	2	3	2	**2,4**	**1,80**	moyen	moyen
Les mystères du bonnet rouge	1	3	1	1	**1,4**	1	2	1	**1,3**	3	2	2	3	2	**2,4**	**1,70**	moyen	moyen
L'affaire Brierre : le massacre des innocents	1	3	1	1	**1,4**	2	2	2	**1,6**	3	1	2	3	2	**2,2**	**1,73**	moyen	moyen
L'archevêque assassiné. Papavoine le tueur d'e	1	3	1	1	**1,4**	2	2	2	**1,6**	3	2	2	3	2	**2,4**	**1,80**	moyen	moyen
Code théâtral, physiologie des théâtres, manue	1	3	1	1	**1,4**	2	2	3	**1,6**	3	3	2	0	3	**1,8**	**1,60**	moyen	difficile
Souvenirs de la vie littéraire	2	2	1	3	**2**	3	3	3	**2,4**	3	3	0	0	2	**1,1**	**1,83**	moyen	moyen
Qu'est-ce que l'occultisme ? : étude philosophi	1	3	1	1	**1,4**	2	2	3	**1,6**	3	2	2	3	2	**2,4**	**1,80**	moyen	moyen
Nouveau manuel complet du physionomiste des	1	2	1	2	**1,4**	2	2	3	**1,6**	3	2	0	0	3	**1**	**1,33**	difficile	difficile
Histoire philosophique et politique de l'occulte :	2	3	1	1	**1,8**	3	3	3	**2,4**	3	2	0	0	2	**0,9**	**1,70**	moyen	moyen
La science occulte : magie pratique, révélation	1	3	1	1	**1,4**	3	3	3	**2,4**	3	2	0	0	2	**0,9**	**1,57**	moyen	difficile
Mystères des sciences occultes : physiognom	2	3	1	1	**1,8**	2	3	2	**2,1**	3	2	0	0	2	**0,9**	**1,60**	moyen	difficile
La clef des grands mystères : suivant Hénoch,	2	3	1	1	**1,8**	2	3	2	**2,1**	3	2	0	0	2	**0,9**	**1,60**	moyen	difficile
Les XXII lames hermétiques du tarot divinatoire,	2	3	1	1	**1,8**	3	3	3	**2,4**	3	3	3	0	2	**2**	**2,07**	moyen	moyen
Traité des voitures pour servir de supplément au	2	2	1	2	**1,8**	2	2	3	**1,6**	3	3	2	0	3	**1,8**	**1,73**	moyen	moyen
Les crimes de la duchesse / par Odysse Barot	2	2	1	2	**1,8**	3	3	3	**2,4**	1	2	0	0	2	**0,7**	**1,63**	moyen	difficile
La divination et la science des présages chez l	1	2	1	3	**1,6**	1,5	3	3	**1,95**	3	3	2	0	2	**1,7**	**1,75**	moyen	moyen
De la Musicothérapie. La musique associée à l	2	2	1	3	**2**	2	2	2	**1,6**	3	2	3	3	2	**2,7**	**2,10**	moyen	moyen
Les véritables moyens pour forcer l'amour : plus	2	3	1	1	**1,8**	2	2	2	**1,6**	3	3	2	0	2	**1,7**	**1,70**	moyen	moyen
Le Premier livre illustré de mes petits enfants,	3	2	3	3	**2,8**	3	3	3	**2,4**	2	3	3	3	2	**2,8**	**2,67**	facile	facile
Les aventures d'un apprenti parisien : le tour du	2	2	1	2	**1,8**	2	2	2	**1,6**	2	2	3	3	2	**2,6**	**2,00**	moyen	moyen
Manuel du cuisinier et de la cuisinière (4ème é	1	3	1	1	**1,4**	0	1	2	**0,5**	3	2	0	0	3	**1**	**0,97**	difficile	difficile
L'assassinat du courrier de Lyon / Arthur Bernè	1	3	1	1	**1,4**	1	2	1	**1,3**	3	1	2	3	2	**2,2**	**1,63**	moyen	difficile
Les secrets de Bolo dévoilés	1	3	1	1	**1,4**	1	2	1	**1,3**	3	1	2	3	2	**2,2**	**1,63**	moyen	difficile
Le vampire de Dusseldorf	1	3	1	1	**1,4**	2	2	2	**1,6**	3	2	2	3	2	**2,4**	**1,80**	moyen	moyen
La femme Weber : l'ogresse de la Goutte d'Or	1	3	1	1	**1,4**	2	2	1	**1,6**	3	1	2	3	2	**2,2**	**1,73**	moyen	moyen
Manuel simplifié de l'accordeur, ou l'Art d'accor	3	3	3	3	**3**	3	3	3	**2,4**	3	3	3	3	3	**3**	**2,80**	facile	facile
Code de la conversation, manuel complet du la	2	3	2	1	**2**	3	2	3	**1,9**	3	3	2	0	3	**1,8**	**1,90**	moyen	moyen
J. François-Oswald. L'Appel aux armes	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
Henry Frichet. La Bataille de l'Yser-Dixmude. L	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
G. Le Rouge. Le Carnet d'un reporter	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
Georges Spitzmuller. Épisodes de la vie d'un 4	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
Manuels Roret. Nouvelle théorie des sapeurs-p	3	3	3	3	**3**	3	3	3	**2,4**	3	3	3	3	2	**2,9**	**2,77**	facile	facile
Nouveau manuel complet du charcutier... par M	3	3	3	3	**3**	3	3	3	**2,4**	3	2	0	0	2	**0,9**	**2,10**	moyen	moyen
A. de Laporte. A Beauséjour : les téléphonistes	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
Julie Crémieux. Souvenirs d'une infirmière	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
C. Altam. Un Parisien à Salonique	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
Henry Frichet. La Bataille de l'Ourcq	3	3	3	3	**3**	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile
A. de Laporte. Aux Éparges : Zizi, agent de liai	3	3	3	3	3	3	3	3	**2,4**	3	2	3	3	2	**2,7**	**2,70**	facile	facile