

ANALYSER LA PRESSE ANCIENNE AVEC LE PROGICIEL *PhPress* : LE TRAITEMENT NUMÉRIQUE DES FAITS DIVERS DE *L'ÉCLAIREUR DE NICE*, 1928-1929

Matthieu PEREZ
CMMC, Université de Nice - Sophia Antipolis

SOMMAIRE

1. Les raisons d'être de *PhPress*
 - 1.1. Brève historiographie du fait divers
 - 1.2. Les obstacles techniques limitant les champs de recherche de ce domaine
 - 1.3. Les solutions théoriques pour y remédier
 - 1.4. La mise en application de ces solutions et la construction de *PhPress*
2. La constitution du corpus documentaire
 - 2.1. L'acquisition numérique des documents depuis les sources originales ou microfilmées
 - 2.2. Repérage et sélection des articles de faits divers
 - 2.3. L'étiquetage simple des articles : titre, thématique principale, géographie, morphologie
 - 2.4. L'étiquetage avancé des articles : thématiques secondaires, personnages, relations entre articles, relations entre personnages, géographie fine
3. Exploitation de la base et interprétation des résultats
 - 3.1. Axe morphologique
 - 3.2. Axe géographique
 - 3.3. Taxinomies
 - 3.4. Exploitation avancée des données

Résumé : *PhPress* est un progiciel de Gestion Électronique de Documentation conçu pour assister l'historien dans le traitement de la presse ancienne. Développé dans un esprit Open-Source, il s'agit d'un outil puissant permettant la numérisation de vastes fonds de presse quotidienne, la consultation aisée de ces fonds, ainsi que la construction de bases analytiques destinées à l'étude du contenu et de la morphologie des articles. Nous avons construit cet outil en fonction des besoins que nous avons définis au cours de l'élaboration de notre projet de thèse : « La chronique des faits divers dans la presse niçoise des années de crise : 1929-1939 » (thèse en cours de réalisation, sous la direction de M. Ralph SCHOR).

Nous chercherons dans un premier temps à présenter les raisons de la construction de ce logiciel, avant d'en aborder le fonctionnement concret dans le cadre d'une étape de notre travail de recherche, l'analyse de la chronique des suicides dans le journal régional *L'Éclaireur de Nice* de l'année 1928, constituée de deux temps : d'une part la constitution du corpus numérique, d'autre part l'exploitation de ce corpus.

1. Les raisons d'être de *PhPress*

La construction d'un logiciel dans le cadre d'un travail de recherche en Sciences Humaines n'est pas très fréquente et il peut sembler nécessaire d'expliquer ce qui a motivé ce travail très lourd et assez inhabituel.

1.1. Brève historiographie du fait divers

Le fait divers est une rubrique de presse constituée de récits d'événements anormaux prenant place dans le cadre de la vie quotidienne : crimes, délits, accidents, anomalies naturelles, faits curieux de toute sorte. Ce sont des récits généralement considérés comme insignifiants, vulgaires, médiocres. Ils occupent pourtant dans la presse une place considérable, particulièrement dans la presse antérieure aux années 1960, et entretiennent avec la littérature une relation très forte : beaucoup d'écrivains ont débuté leur carrière comme fait-diversiers, beaucoup de livres importants ont pour origine un fait divers.

C'est dans le champ littéraire que les bases de l'analyse du fait divers vont être posées : Roland Barthes en définit la structure narrative en 1964. Puis Georges Auclair l'observe sous un angle anthropologique en 1970.

On a longtemps considéré ce genre de récit comme non historique : le fait divers apparaissait comme un objet littéraire ou anthropologique, mais ne trouvait place que dans une « petite histoire » anecdotique opposée à la « grande histoire ». Observé sous l'angle de l'anecdote, le fait divers apparaissait avant tout comme un divertissement, et beaucoup de titres consacrés à ce sujet sont d'abord ludiques ; l'historien utilisait ce matériau avant tout comme simple illustration.

Avec Dominique Kalifa et ses recherches sur le récit de crime à la belle-époque, qui présente le fait divers et la littérature populaire comme un ensemble de récits normalisateurs, l'appréhension du fait divers par l'historien prend un nouvel aspect.

Deux thèses entièrement consacrées au fait divers vont immédiatement suivre, avec Marine M'Sili et Anne-Claude Ambroise-Rendu, qui abordent le fait divers à travers des méthodes nouvelles : le comptage des articles, des références géographiques qu'ils contiennent, la ventilation des articles en fonction de leur contenu, prennent une importance centrale dans le traitement des sources, dont l'interprétation se fait largement au moyen de méthodes statistiques. Ces nouvelles méthodes correspondent à une prise en compte du récit de fait divers comme un objet historique à part entière, où le contenu, la répétition, la sur ou sous-représentation de certains thèmes, vont être porteurs de sens.

Le travail que nous réalisons actuellement vise à étudier, dans le contexte très particulier de la Riviera durant les dix années précédant la seconde guerre mondiale, le comportement de la chronique des faits divers de deux quotidiens locaux, *L'Éclaireur de Nice* et *Le Petit Niçois*, représentant deux tendances politiques opposées.

1.2. Les obstacles techniques limitant les champs de recherche de ce domaine

La rubrique des faits divers occupe généralement une très grande place dans la presse ancienne, et on en trouve habituellement 30 à 60 articles par numéro dans les quotidiens généralistes. Cela représente très vite, même sur une période relativement courte, une masse documentaire considérable, dont la simple manipulation devient problématique.

Tant que le traitement des documents reste purement manuel, cette surabondance de matière, indissociable de cette nouvelle approche méthodologique plaçant le fait divers et son contenu au centre de la problématique, oblige les chercheurs à procéder à des sondages très dilués sur leur corpus, et l'on perd ainsi énormément d'information : par exemple, le rythme saisonnier de la chronique va être masqué par un sondage ne portant que sur une courte partie de l'année. Par ailleurs, un simple comptage des articles, s'il permet l'élaboration de tableaux statistiques utiles, ne permet pas de les hiérarchiser et de faire la différence entre une information importante et une simple dépêche.

Les méthodes d'analyse morphologique de la presse quotidienne définies par Jacques Kayser permettraient de lever ce genre d'ambiguïtés, mais elles sont difficilement utilisables sur un corpus important : mesurer la surface de quelques articles dans quelques numéros est aisé, le faire pour plusieurs centaines ou plusieurs milliers d'articles est impossible, du moins dans des délais acceptables.

Enfin, il est très difficile de revenir en arrière et de retrouver un article particulier ou mal saisi lorsque l'on travaille de façon traditionnelle sur un corpus de presse quotidienne : copier tous les articles étudiés serait trop coûteux, trop long, et poserait le problème de la gestion d'une telle masse de documents ; la manipulation des microfilms ou des liasses d'archive est suffisamment compliquée pour que l'on ne revienne à un document déjà consulté qu'en cas de très grande nécessité, et que l'on ne travaille habituellement qu'à partir de notes, prises parfois un peu vite. Ces difficultés, observées concrètement lors de la réalisation de notre mémoire de maîtrise, nous semblaient très handicapantes, et nous avons recherché une solution technique pour les dépasser.

1.3. Les solutions théoriques pour y remédier

Nous avons cherché le moyen d'optimiser autant que possible le travail du chercheur, de façon à augmenter le volume de documents et de données manipulés dans un temps aussi court que possible, afin de rendre possible un travail sur le fait divers mettant en œuvre un corpus vaste, accroissant la précision de la recherche et la fiabilité des statistiques obtenues. Il fallait donc un logiciel adapté. Les contacts que nous avons eus avec les services informatiques universitaires et avec différentes entreprises professionnelles de la Gestion Électronique de Documentation se sont

vite avérés décevants : si de nombreux systèmes de gestion de documentation existaient, aucun ne permettait de faire efficacement le travail que nous projetions.

Nous avons donc déterminé un cahier des charges pour une application informatique, qui devrait contenir une base de données contenant l'ensemble de l'analyse de contenu appliquée aux faits divers, optimisant ainsi le travail statistique classique réalisé sur cette rubrique, et qui permettrait également de mener une analyse morphologique des articles, ainsi que de les conserver sous forme électronique, pour avoir la possibilité de les retrouver et de les consulter aisément.

Dans tous les cas, il nous semblait primordial de réduire autant que possible les gestes à accomplir pour utiliser le logiciel, et d'optimiser au maximum la productivité du système.

D'autre part, il nous a semblé utile de chercher à réaliser un logiciel qui serait un logiciel libre, pour que son développement puisse éventuellement se poursuivre dans le cadre de la communauté Open-Source, pour que sa diffusion puisse se faire sans contrainte, et surtout pour en rendre visibles les mécanismes : en théorie, chacun peut ainsi vérifier que le logiciel fonctionne réellement et ne donne pas de résultat erroné ; d'un point de vue pratique, le choix du développement Open-Source permet également de faire appel à de nombreux logiciels libres existants, et de les inclure dans le fonctionnement du système – ce qui ne serait pas possible dans une logique de développement propriétaire.

Nous voulions également que le logiciel soit aussi modulaire que possible, et puisse servir à d'autres recherches que la nôtre.

Enfin, nous voulions garantir une certaine interopérabilité, et faire en sorte que le système soit utilisable depuis des ordinateurs fonctionnant avec des applications Microsoft, mais également sous Linux et Mac OS.

À partir de ce cahier des charges, des choix techniques précis se sont imposés de façon assez automatique.

1.4. La mise en application de ces solutions et la construction de *PhPress*

Parmi les diverses solutions techniques disponibles, les différents langages de programmation, le choix d'un système basé sur PHP et MySQL s'est rapidement imposé : il s'agit de techniques Open-Source, très populaires et donc bien documentées, d'un abord relativement facile ; PHP est un préprocesseur Hypertexte, qui génère automatiquement des pages Web, et qui dispose de nombreuses fonctions de traitement graphique et logique ; il permet de construire facilement des interfaces vers un système de gestion de bases de données relationnelles tel que MySQL, un logiciel libre, assez facile d'accès et très répandu.

Nous avons donc appris les bases de la programmation PHP, et commencé la construction du logiciel – construction encore inachevée, puisque bien que déjà utilisable, *PhPress* ne dispose pas encore de toutes les fonctionnalités que nous avons prévues.

Techniquement, *PhPress* fonctionne comme un site Web dynamique ; il met en œuvre un ordinateur employé comme serveur, qui doit obligatoirement fonctionner sous Linux, qui contient la base de données et les images représentant les pages de journal numérisées, et un ordinateur client, utilisé pour se connecter au serveur, qui doit, pour présenter de bons résultats, être un Macintosh disposant de Mac OSX et du navigateur Firefox. L'interopérabilité que nous recherchions n'est pas encore atteinte, mais reste un objectif à long terme – notre priorité étant de réaliser notre propre travail de recherche.

2. La constitution du corpus documentaire

2.1. L'acquisition numérique des documents depuis les sources originales ou microfilmées

Au début de notre travail, la BNF ne prévoyait pas encore de numériser ses collections de journaux anciens, et les archives départementales des Alpes-Maritimes ne pouvaient pas avancer de date ni donner de détail technique concernant leur campagne de numérisation de la presse quotidienne locale. Nous avons donc choisi, en accord avec la direction des Archives Départementales, de numériser nous-mêmes, à partir d'un lecteur-numériseur de microfilms, les collections dont nous avons besoin, en commençant par l'année 1928.

Nous avons donc équipé *PhPress* d'un dispositif permettant d'indexer manuellement des images au format TIFF (le seul format d'image possible avec le matériel dont nous disposions) au fur et à mesure de la numérisation d'un rouleau de microfilm, et nous avons ainsi numérisé la collection complète du principal quotidien régional de la période, *L'Éclaireur de Nice*, entre 1928 et 1932, soit

12 214 pages ; nous avons construit des outils permettant de gérer la sauvegarde sur CD-ROM de ces documents et de les convertir dans un format manipulable par PHP, le format PNG.

Depuis, les Archives Départementales des Alpes-Maritimes ont achevé une tranche importante de leur propre campagne de numérisation de leurs collections, qui incluent *L'Éclaireur de Nice* et *Le Petit Niçois* ; ces journaux ont été numérisés depuis les pages originales, ce qui permet une qualité bien supérieure à celle obtenue par numérisation des microfilms. Ces documents ont été mis en ligne et sont accessibles depuis le site des Archives Départementales, mais avec un temps de réponse très long du serveur et une interface qui ne se prête pas à une exploitation automatique, ce qui rend impossible l'exploitation systématique de ces pages ; la direction des archives a accepté de nous transmettre une copie de ces documents, mais la constitution de cette copie semble poser des problèmes et nous restons pour l'instant dans l'expectative. Nous ignorons encore les spécificités techniques de ces images, et nous ne pourrions donc construire un module permettant de les intégrer à *PhPress* que lorsque cette situation sera débloquée.

2.2. Repérage et sélection des articles de faits divers

Les images numériques représentant les pages de journal doivent, avant d'être utilisées, subir un étalonnage permettant de connaître leur surface réelle, de façon à compenser les différences d'échelle dues à la focale optique du numériseur de microfilm. Nous avons ainsi réalisé une mesure de la surface imprimée réelle des pages de *L'Éclaireur de Nice* pour la période qui nous concerne – période où le format du journal reste assez stable – qui permet de convertir approximativement en millimètres les surfaces en pixels des images informatiques.

Pour chaque page, l'utilisateur doit donc désigner d'un clic les quatre points extrêmes de la surface imprimée de la page affichée ; à partir des coordonnées de ces points, l'ordinateur calculera la surface relative de chaque article qui sera sélectionné dans cette page, sélection qui se fait également d'un clic sur les quatre points extrêmes de la surface choisie. À partir des coordonnées de ces points, l'ordinateur calculera la surface de chaque article et estimera sa position dans la page de journal. La sélection des articles se fait également par sélection à la souris des quatre coins du bloc de texte ou d'illustration. Un masque de couleur transparente est appliqué sur le document de façon à rendre visible le travail réalisé, et de distinguer les différents types d'articles présents sur la page.

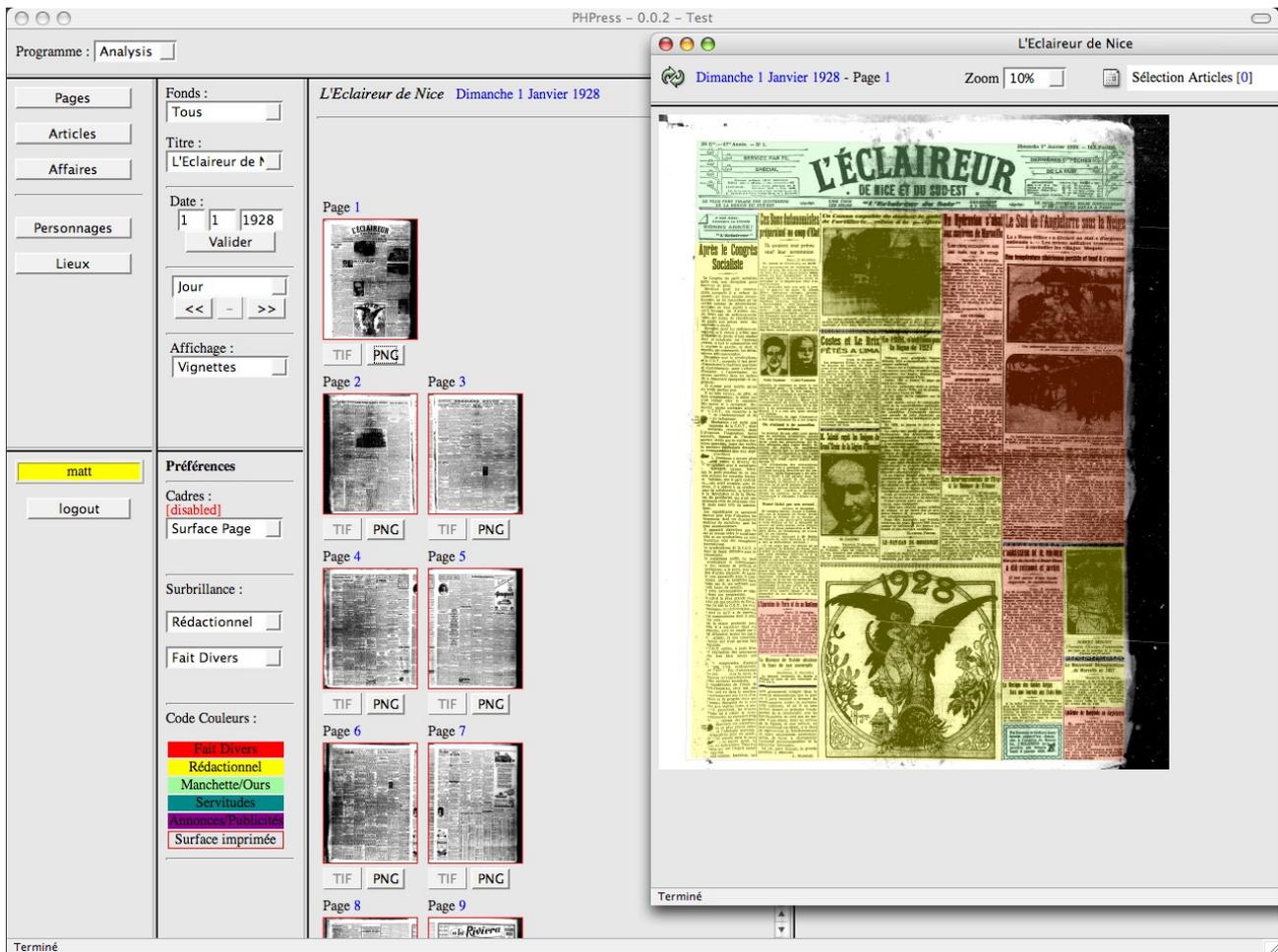


Illustration 1 : exemple de page de journal traitée avec PhPress

2.3. L'étiquetage simple des articles : titre, thématique principale, géographie, morphologie

Une fois un article sélectionné, un masque apparaît automatiquement et demande à l'utilisateur de saisir les informations essentielles permettant d'indexer ce document : son titre, sa nature (article rédactionnel, rubrique de servitudes, publicité, élément de titre), sa typologie (article de fait divers, de politique, etc.) ainsi que la sous-catégorie dans laquelle on peut le placer : on pourra ainsi ventiler les faits divers entre récits de suicide, récits de crime, etc., et analyser séparément ces différents genres de faits divers.

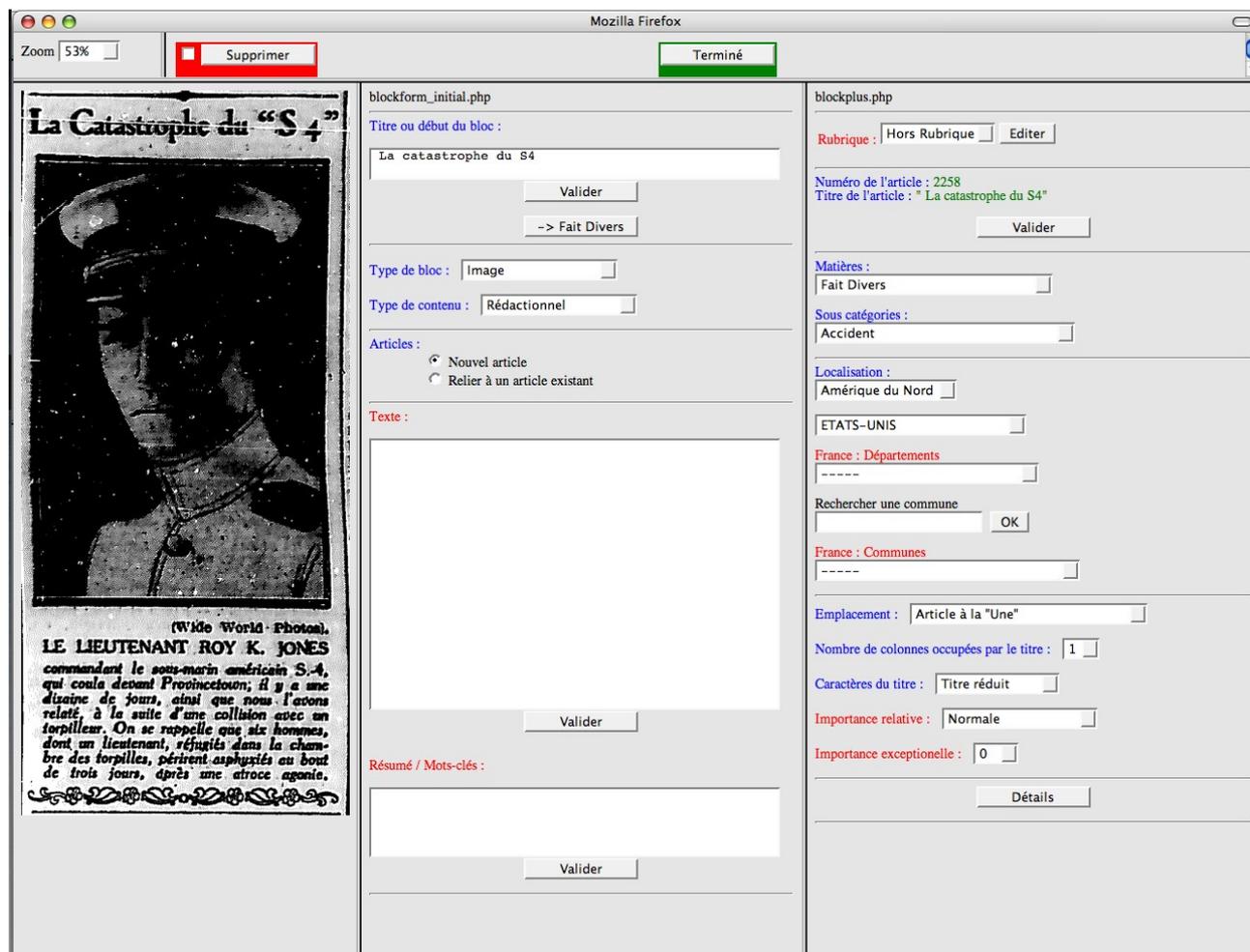


Illustration 2 : un exemple de masque d'indexation des articles

On peut relier chaque article à une indication géographique plus ou moins précise : continent, pays, et, pour la France, département et commune. Les données géographiques sont calquées sur les bases de données de l'INSEE.

Le système prévoit également un formulaire permettant de calculer un indice de mise en valeur inspiré par la méthode de Jacques Kayser et partiellement automatisée, qui permet de hiérarchiser les articles en fonction de leur aspect graphique. Cet indice est complémentaire d'une autre information essentielle calculée de façon automatique par le logiciel, la surface de l'article – exprimée en pourcentage de la surface imprimée de la page :



Illustration 3 : affichage de la surface relative de l'article, de l'indice de mise en valeur et de la position de l'article

L'indication de position de l'article est un chiffre de 1 à 9 correspondant à neuf secteurs de la page de journal, calculé automatiquement à partir des coordonnées du coin supérieur gauche de l'article.

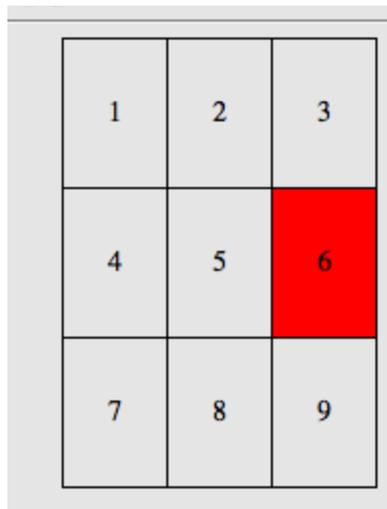


Illustration 4 : repérage de la position de l'article dans la page

Une fois indexé, l'article est accessible depuis l'affichage analytique correspondant au numéro du journal :

Programme : Analysis Language : FR

Fonds : Tous
Titre : L'Eclaireur de N
Date : 2 1 1928
Affichage : Analyse

L'Eclaireur de Nice Lundi 2 Janvier 1928

Page 1 :

Article	Titre	Matières
193	l'oranie de nouveau sous l'eau Hors Rubrique ... - ... - ALGERIE	Fait Divers Météors 14.4334% [47-5]
2258	La catastrophe du S4 Hors Rubrique ... - ... - ETATS-UNIS	Fait Divers Accident 3.1634% [38-6]

Page 2 :

Article	Titre	Matières
194	La fête d'un bataillon de Highlanders dégénère en graves désordres Hors Rubrique ... - ... - ROYAUME-UNI	Fait Divers Violences diverses 2.2082% [4-7]
195	marié seize fois en cinq mois Hors Rubrique ... - ... - ETATS-UNIS	Fait Divers Déviances 2.3300% [4-7]

Page 3 :

Article	Titre	Matières
196	le mauvais temps a recommencé Hors Rubrique Paris - Paris - FRANCE	Fait Divers Météors 2.0000% [0-2]
197	la tempête souffle dans la manche Hors Rubrique Cherbourg-Octeville - Manche - FRANCE	Fait Divers Météors 1.3955% [2-2]
198	un ouragan fait 19 victimes à Chicago Hors Rubrique ... - ... - ETATS-UNIS	Fait Divers Météors 0.9023% [2-5]

Dernier article étudié :
L'Eclaireur de Nice - Lundi 2 Janvier 1928 - page 1
Titre de l'article : " La catastrophe du S4"
[Fait Divers]

Numéro en cours :
Lundi 2 Janvier 1928
Surface imprimée : 8/8 pages - 1840000/1840000 mm²
Nombre Articles : 54 - Fait Divers : 54
Rédactionnel : 54
Servitudes : 0
Annonces/Publicités : 0
Manchette/Ours : 0

Méthode d'analyse :
64 numéro(s) avant la prochaine analyse complète
1 numéro(s) avant la prochaine analyse partielle
Analyse complète : 1 numéro sur 65
Analyse partielle : 1 numéro sur 1

Marqueur de ce numéro :
 Analyse complète
 Analyse partielle
 Passé sans analyse

Illustration 5 : affichage analytique du numéro du 2 janvier 1928 de L'Éclaireur de Nice

Les articles, après ce premier marquage, peuvent déjà être triés et comptés en fonction de leur position, de leur taille, de leur typologie, des lieux géographiques mentionnés. Un étiquetage plus fin est également possible.

2.4. L'étiquetage avancé des articles

Les articles s'affichent dans une fenêtre mettant en regard trois cadres, d'abord l'article lui-même, le cadre central contenant un résumé des informations contenues dans l'article, le cadre de droite permettant l'édition de ces informations :

The screenshot displays the PHPRESS interface for editing an article. At the top, the browser window title is 'Double noyade dans le Var'. The address bar shows the URL 'Lundi 2 Janvier 1928, L'Eclaireur de Nice, page 4 [Art. 215] - Surface/page : 14.2122% - ind.9 / pos.2'. The main content area is divided into three sections:

- Left Panel (Bloc 248):** Contains the article's title 'Double Noyade dans le Var' and a sub-headline 'Une femme se suicide; son compagne se noie en tentant de la sauver'. Below this is the start of the article text, including the date 'Lundi 2 Janvier 1928' and the newspaper name 'L'Eclaireur de Nice'.
- Center Panel (article_main.php - 215):** A table of metadata with fields for Rubrique (Faits-Divers), Titre ('Double noyade dans le Var'), Matières (Faits Divers => Suicide), Lieux (06088 - Nice - Alpes-Maritimes - FRANCE), Evenement (noyade), Raison (l'amour), Issue (plusieurs morts), Thème (Famille désunie), Dossiers (Nomades, gens du voyage), and Affaire.
- Right Panel (article_plus.php - 215):** A list of categories and sub-categories for tagging the article. The 'Suicide' category is selected, and a dropdown menu is open showing sub-categories like 'Suicide', 'Accident', 'Vol', 'Meurtre', 'Incendie', etc.

Illustration 6 : exemple d'affichage d'un article et édition des marqueurs qui le caractérisent

Nous pouvons ainsi ajouter à la description sommaire de l'article réalisée lors de sa sélection des marqueurs plus fins, tels la nature précise de l'événement, les motifs des acteurs, l'issue, le thème principal de l'article. Une fonction permettant d'insérer l'article dans une cartographie précise, tenant compte du nom des rues et permettant de les repérer sur un plan, est en cours d'élaboration.

Il est possible de construire un certain nombre de dossiers, et de relier ainsi les articles présentant des caractéristiques communes, déterminées par le chercheur.

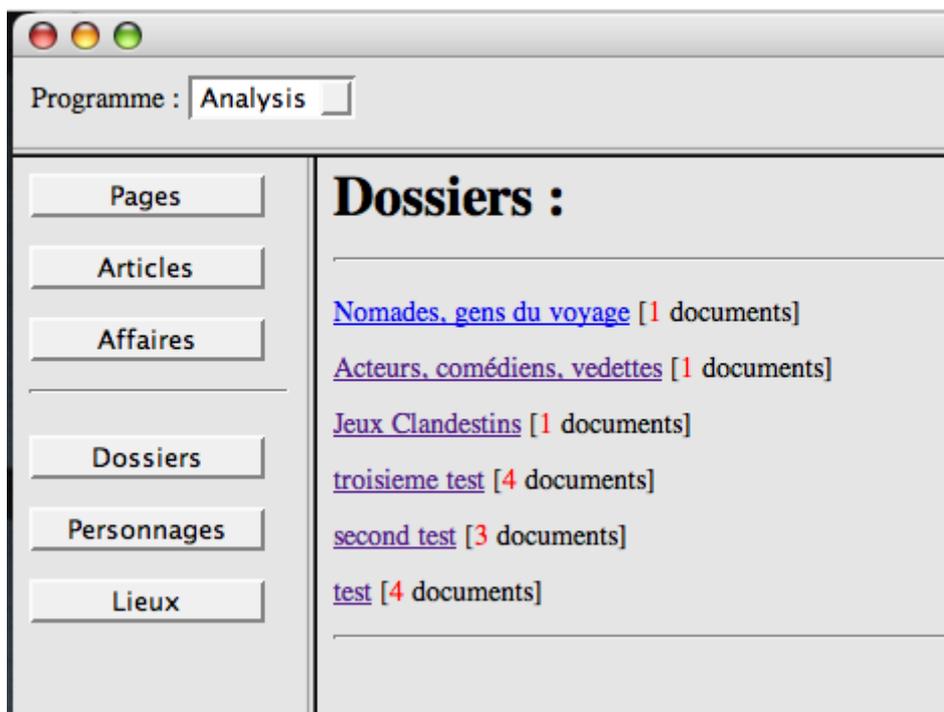


Illustration 7 : liste de dossiers

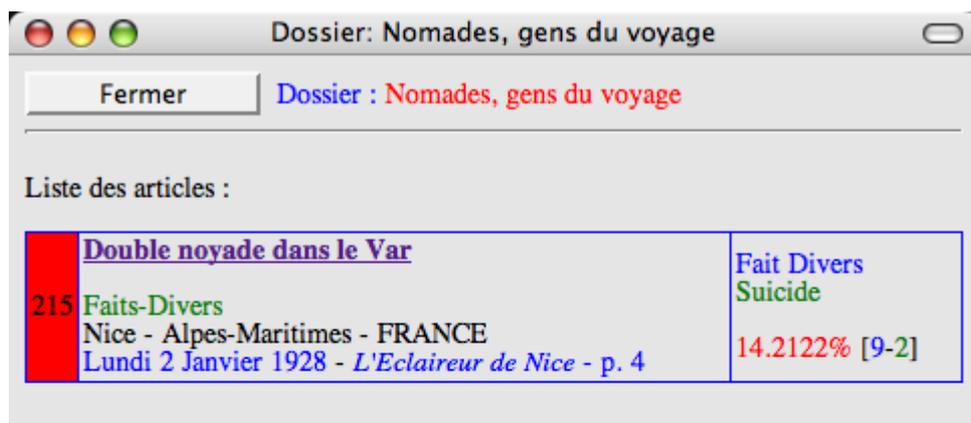


Illustration 8 : un exemple de dossier

La base de données comportera prochainement une table permettant de suivre les articles en tant qu'affaires, c'est-à-dire de repérer les articles se suivant et concernant un même événement, ainsi que les articles équivalents publiés par plusieurs journaux. Une autre table permettra de constituer un fichier des personnes apparaissant dans les articles, et de noter les actions et les relations entre ces personnes. Cependant, nous rencontrons des difficultés techniques dans la réalisation de ces deux éléments, dont le développement a pris beaucoup de retard. D'une façon générale, l'étiquetage avancé des articles est encore en période de test, seul l'étiquetage simple ayant été utilisé sur un volume conséquent de documents.

3. Exploitation de la base et interprétation des résultats

Nous avons étiqueté 2 468 articles de faits divers parus dans *L'Éclaireur de Nice* durant l'année 1928, dont 217 articles représentant des suicides. À partir de cette documentation réduite, nous avons pu faire un test de la productivité de notre système et des contraintes pesant sur son utilisation, qui a entraîné un certain nombre de modifications dans le logiciel. Les routines permettant l'exploitation de la base de données restent pour l'instant très sommaires et manquent de maturité, mais permettent tout de même d'obtenir certains résultats.

3.1. Axe morphologique

La prise en compte de la morphologie des articles est un point très important de notre système. Pour le moment, l'exploitation reste assez limitée et se fait au travers d'un formulaire permettant de trier les articles par thème, par date, par surface et par indice de mise en valeur :

The screenshot shows the PHPRESS application interface. On the left, there is a sidebar with navigation buttons: Pages, Articles, Affaires, Dossiers, Personnages, and Lieux. The main area is divided into two panes. The left pane, titled 'ana_mode_art_search.php', contains search filters: 'Recherche rapide' with fields for 'N°', 'Article', and 'Valider'; 'Recherche par éléments de la base articles' with a dropdown for 'L'Eclairer de Nice'; 'Trier par' with radio buttons for 'Date', 'Mise en valeur', 'Surface', and 'Ordre Inverse' (checked); and 'Afficher selon statut' with radio buttons for 'Tous', 'A traiter', 'Traité B (1)', and 'Traité C (2)'. The right pane, titled 'ana_mode_art_result.php', displays search results. At the top, it shows 'Indice Surface : 115 - 88 résultat(s) - page 1/9'. Below is a table of results:

N°	Titre	Thème	Indice	Surface
215	Double noyade dans le Var	Faits-Divers Suicide	14.2122%	[9-2]
221	les désespérés	Faits-Divers Suicide	1.3325%	[5-3]
222	dimanche matin également...	Faits-Divers Suicide	1.2161%	[0-3]
260	Un vieillard de 99 ans se suicide	Faits-Divers Suicide	0.6346%	[2-3]
269	on découvre à la bocca un nové dont l'identité est inconnue	Faits-Divers Suicide	2.8226%	[6-3]
273	la double noyade du var [suite de l'article de la veille]	Faits-Divers Suicide	2.4231%	[4-5]
285	clauda france s'est suicidée à la suite de chagrins intimes	Faits-Divers Suicide	1.0536%	[4-2]
305	on a identifié le nové de la bocca	Faits-Divers Suicide	0.9762%	[4-5]
331	Une des filles du général russe...	Faits-Divers Suicide	0.2489%	[0-6]
337	Il voulait se suicider... mais une crise cardiaque le jette inanimé sur le sol	Faits-Divers Suicide	2.0141%	[4-5]

Illustration 9 : recherche d'articles de suicide en fonction de leur date de publication

Le système, en fonction de ces critères de recherche, fournit une estimation de la surface réelle cumulée des articles répondant aux critères ; nous avons ainsi pu élaborer des tableaux présentant l'évolution sur un an des récits de suicide dans *L'Éclairer de Nice*.

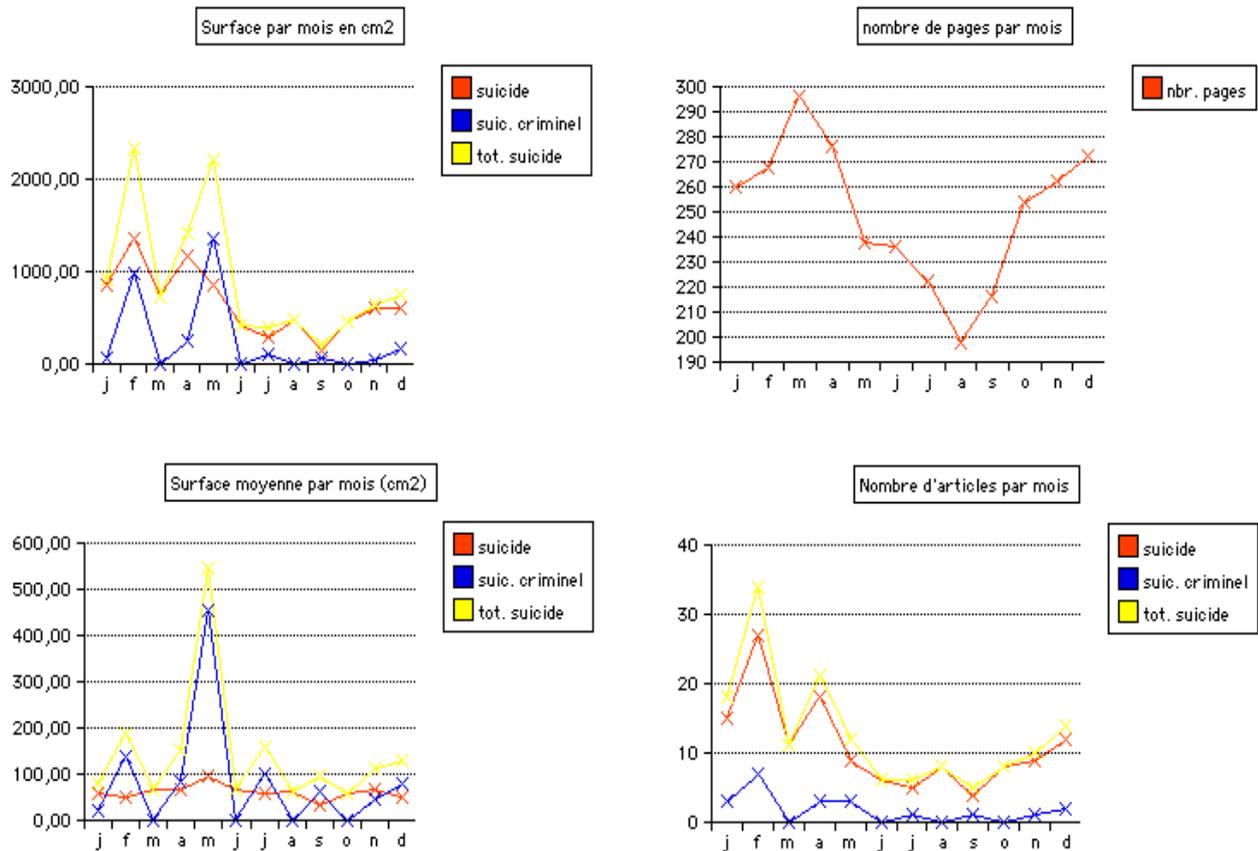


Illustration 10 : évolution des récits de suicide au cours de l'année 1928 dans L'Éclaireur de Nice

L'ensemble du système reste très sommaire et appelle de nouveaux développements, qui sont en cours de réalisation.

3.2. Axe géographique

Nous n'avons pas encore achevé les scripts permettant d'interroger la base de données à partir de critères strictement géographiques, et nous n'avons donc pas encore pu utiliser pleinement cette information. Les données géographiques concernant les articles sont libellées dans les mêmes normes que celles de l'INSEE. Notre objectif à court terme est de permettre un tri simple des articles à partir de critères géographiques ; dans un deuxième temps, la représentation des données sous forme de carte est envisagée, mais nous manquons de temps pour réaliser le développement de ces applications.

Une géographie fine, permettant d'enregistrer le nom des rues où se déroulent les événements, est déjà en place mais n'a pas encore été utilisée à grande échelle.

L'inscription dans l'espace géographique des récits de fait divers est un point essentiel de notre travail de recherche, nous sommes donc très mobilisé par cette question.

3.3. Taxinomies

L'étiquetage des articles se fait avant tout par l'insertion des articles dans une classification : classement par catégories d'articles (politique, faits divers, etc.) et de là en sous-catégories (crime, suicide, sous-catégories de l'ensemble fait divers, ...). Mais concrètement, les articles de presse sont parfois difficiles à classer dans une catégorie trop précise ; la multiplication des cas particuliers ou limites nous a dans un premier temps poussé à multiplier les sous-catégories rares, avant de développer des marqueurs supplémentaires permettant de limiter les rubriques de classement principal, sans perdre les particularités des articles. L'utilisation à grande échelle de *PhPress* nous a poussé à modifier complètement la liste des sous-catégories de faits divers que nous avons établie à l'avance, et qui ne correspondait pas suffisamment à la réalité des articles. Les catégories permettant de marquer les autres articles, que nous avons construites en nous

fiant un peu servilement à une liste de rubriques présentée par Jacques Kayser, se sont révélées pratiquement inutilisables, et doivent être complètement révisées.

3.4. Exploitation avancée des données

PhPress est un système dont le développement est en cours, et qui souffre donc de nombreuses lacunes ; l'exploitation des données reste particulièrement en retard sur d'autres fonctions complexes, dont l'élaboration a pris beaucoup de temps. Nous ne pouvons pour l'instant présenter que des fonctionnalités qui sont en cours d'élaboration ou de conception.

Le principal axe de développement du logiciel est le traitement des séries d'articles, des affaires, qui permettra de déterminer comment des journaux concurrents traitent des mêmes sujets ; il s'agit là d'un point particulièrement important, mais également particulièrement complexe, que nous sommes actuellement en train de mettre au point.

D'autre part, le traitement des personnages représentés fournira des clés importantes, mais demande une réflexion technique importante qui n'est pas achevée.

En conclusion, *PhPress* est un logiciel qui reste inachevé, et qui évolue au cours de l'utilisation que nous en faisons ; il est clair qu'il n'est pas encore pleinement fonctionnel, et qu'il n'est pas utilisable sans un important apprentissage. Construit par un historien, dans un but précis et entièrement déterminé par nos objectifs de recherche, il remplit son rôle essentiel correctement et représente un outil augmentant considérablement la puissance de travail de l'historien. Tant que son utilisateur reste son concepteur, l'outil se plie naturellement aux impératifs du chercheur ; mais, publié, il sera certainement difficilement maîtrisable pour les utilisateurs, même de bon niveau ; il reste donc un très grand travail de développement à réaliser, qui devra viser à rendre l'ensemble du système de *PhPress* compréhensible et aisément modifiable par les utilisateurs.

BIBLIOGRAPHIE

AMBROISE-RENDU, A.-C. 1997. *Les faits divers dans la presse française de la fin du XIXe siècle. Étude de la mise en récit d'une réalité quotidienne (1870-1910)*, Thèse de doctorat, Paris 1 Sorbonne, 765 p.

AUCLAIR, G. 1970. *Le Mana quotidien : structures et fonctions de la chronique des faits divers*, Paris, Anthropos, [réed. 1982, augmentée d'une préface de l'auteur et de l'essai *Le double imaginaire de la modernité dans la vie quotidienne*], 300 p.

BARTHES, R. 1964. Structure du fait divers, *Essais critiques*, Paris, Seuil, pp. 188-197.

KALIFA, D. 1995. *L'encre et le sang : récits de crimes et société à la Belle Époque*, Paris, Fayard, 351 p.

KAYSER, J. 1967. *Le quotidien français*, Paris, Armand Colin, coll. "Cahiers de la fondation nationale des sciences politiques", 169 p.

M'SILI, M. 2000. *Le fait divers en république*, Paris, CNRS Éditions, 311 p.