



www.ichim.org

Les institutions culturelles et le numérique
Cultural institutions and digital technology

École du Louvre
8 - 12 septembre 2003

**EDITION ELECTRONIQUE DE LA RECEPTION DE
NIETZSCHE DES ANNEES 1865 A 1945**

**Daniel Biella, Eva Dyllong, Herbert Kaiser, Wolfram
Luther, Thomas Mittmann - Université de Duisburg-
Essen**

« Acte publié avec le soutien de la Mission de la Recherche et
de la Technologie du Ministère de la Culture et de la Communication »

Résumé

Dans cette note, nous traitons l'édition électronique de la réception allemande des idées et de l'œuvre de Nietzsche. Des textes canoniques de la réception religieuse et philosophique, féministe ou antisémite soigneusement choisis en fonction de leur accessibilité et de leur intérêt, tant scientifique que culturel, ont été répertoriés. Nous expliquons rapidement notre processus intégral de la saisie comprenant la numérisation, la création de schémas mots-clés, l'enrichissement des documents par méta-données, correction et balisage, la transformation en document dans un format de documents standardisé de type « eXtensible Markup Language » XML/HTML, la mise en contexte et la gestion de l'archive. Nous avons placé un point d'intérêt sur la création d'une interface pour piloter la reconnaissance complète ou partielle du texte, ainsi que d'un moteur de recherche permettant la recherche floue, phonétique et par troncation. Plusieurs figures illustrent l'implémentation de ces idées. Afin de perfectionner le processus de sélection, de conversion et de correction des documents, nous avons récemment lancé un projet intitulé « Recherche mobile et répartie de documents dans les archives et bibliothèques ». Il a pour but de développer un scénario comprenant des ordinateurs mobiles intégrés dans un réseau WLAN avec adresses client IP temporaires dans nos bibliothèques, ce qui permet la recherche de documents sur place pilotée en session multimédia par un groupe de spécialistes au départ d'un lieu opportun.

Mots-clés: Edition électronique, réception Nietzscheenne, numérisation, reconnaissance partielle, recherche phonétique, recherche mobile et répartie

Abstract

In this paper we describe our digitization project for the German reception of Nietzsche. Canonical texts in the fields of the religious, philosophic, feministic and anti-Semitic reception were selected depending on their accessibility, scientific and cultural value and were placed into a Web-based repository. We give an overview of our digitization process beginning with the selection, the digital conversion, the creation of keyword-schemes, the meta data extraction and encoding, the document encoding using XML/HTML standards, and the collection management. A focus lies on full and partial optical character recognition, and the develop-

ment and use of search engines supporting a phonetic and wild card search or different degrees of matching in parallel. To optimise the process of selecting and capturing we recently started a funded project entitled “Distributed mobile selection and evaluation of documents in libraries and archives”. It aims at developing a scenario giving researchers with WLAN notebooks access to the libraries’ network via temporary IP addresses. Documents are captured by book scanners or CCD-cameras and uploaded to a document server. The selection is guided by a group of specialists who judge the image quality and relevance of the chosen documents and give recommendations concerning an appropriate conversion process.

Keywords: Electronic edition, reception of Nietzsche, digitization, partial text recognition, phonetic search, distributed mobile document selection

Motivation et intérêt de recherche

L’histoire de la réception de Friedrich Nietzsche a connu un intérêt croissant dans le monde entier, comme en font preuve ouvrages, colloques et conférences. Des chercheurs ont pris en main l’édition chronologique et scientifique de l’ensemble des œuvres et des manuscrits légués par le philosophe, discutent de l’influence de sa pensée sur ses contemporains et leurs descendants en tentant de cataloguer l’immense fond de la littérature réceptionniste parue dans des collections de lettres, journaux et livres. Dans ce contexte, une bibliographie rétrospective et complète pour la période de 1867 à 1998 a été élaborée au sein de la bibliothèque Anna Amalia à Weimar, sous l’égide des chercheurs de la Stiftung « Weimarer Klassik » soutenue par la « Deutsche Forschungsgemeinschaft » DFG (WNB, 2002). Même si cette bibliographie sur la vie, l’œuvre et l’influence du philosophe représente un progrès remarquable, l’accès par ordinateur aux sources, la gestion et l’usage des documents ne sont actuellement pas réglés de façon suffisante, bien que les techniques d’édition électronique et les services bibliothécaires aient connu un développement rapide. Ainsi, les archives de Nietzsche à Weimar hébergent un trésor de textes dont 34000 lettres manuscrites sur l’œuvre et la vie du philosophe pas suffisamment répertorié. D’autres documents sont répartis sur toute l’Europe. Un aspect important qui gêne la numérisation efficace des textes est le fait que dans cette période, les polices « Fraktur » sont utilisées en exclusivité. Chaque imprimerie a employé son jeu de caractères spéci-

que, ce qui rend la reconnaissance optique par scanner extrêmement difficile et demande l'entraînement du logiciel ROC sur chaque spécimen de texte avec un taux d'erreur souvent décevant et inacceptable. En conséquence, la numérisation des années passées s'est souvent restreinte à la période « après-Fraktur » suivie de la procédure plus classique d'un saisissement manuel ou de transformation de texte uniquement en image numérique facsimilé.

Ce fut le point de départ des chercheurs de différentes disciplines de l'université de Duisburg, (y compris l'informatique, les lettres et la linguistique) soutenus dans leur travail par le collège Friedrich Nietzsche à Weimar. Ils se sont réunis sous le nom de « CD Nietzsche » afin de créer des archives de la réception Nietzscheenne des années 1867 à 1945 (Bick, 2003; Biella et al., 2002, 2003; Mittmann, 2001; Pilz, 2003). Un premier effort a concerné les livres, extraits de journaux et textes canoniques de la réception religieuse et philosophique, auxquels se sont ajoutées des œuvres d'auteurs socialistes, féministes ou antisémites. Il s'agissait des travaux d'initiation à la recherche exécutés par les élèves de la deuxième et troisième année d'études en « communication et médias ». Depuis lors, près de 1800 pages de texte soigneusement choisies en fonction de leur accessibilité et de leur intérêt, tant scientifique que culturel, ont été répertoriées, numérisées, enrichies de méta-données structurelles, descriptives et contextuelles et ont été intégrées dans un système informatique permettant l'accès direct à l'aide d'un moteur de recherche proposant la recherche floue, phonétique et par troncation.

Les processus de la numérisation et de la gestion électronique d'archives sont désormais au centre d'une action concertée des informaticiens des archives, des hommes de lettres et des services bibliothécaires visant d'une part la préservation du patrimoine culturel mais aussi l'accès facilité des particuliers en ligne avec des outils standardisés sans perdre de vue les intérêts des auteurs et éditeurs.

L'analyse scientifique d'importantes collections de textes assistée à l'ordinateur permet de reconstruire de façon efficace la vie politique, sociale et culturelle des époques historiques. A partir d'une simple conservation de textes, la numérisation franchit une barre et permet de créer des liens entre textes de différentes provenances et contribue ainsi à la généalogie des idées et des notions.

Nos archives vont être prises en considération par les scientifiques qui se vouent à la reconstruction de l'histoire de la réception et de l'influence de Nietzsche. Notre vision met à leur disposition un corps complet de textes intégraux corrigés munis d'une recherche

phonétique ou floue, permettant ainsi l'analyse qualitative et quantitative des documents sous différentes formes de thèses philosophiques, historiques ou sociologiques.

L'encadrement des projets de numérisation

L'intérêt croissant porté aux nouveaux projets de numérisation doit être apprécié dans le contexte de l'édition de documents imprimés et distribués de façon traditionnelle par les maisons d'édition et les bibliothèques et de la réimpression ou conservation par microfiche des documents face à l'édition numérique ou numérisation rétrospective des documents avec accès par Internet. La plupart des éditeurs ont choisi la double voie en faisant accompagner chaque ouvrage ou périodique d'une version électronique. C'est pourquoi il existe à l'heure actuelle un nombre déjà impressionnant de périodiques distribués par voie électronique. Les budgets des universités et bibliothèques en baisse permanente et la revendication de l'utilisateur d'avoir accès à toutes les publications relevantes à la ligne à partir d'une interface Web standardisée ont fait surgir une vague de projets de numérisation accompagnée d'initiatives en standardisation des méta-données structurelles et administratives. En Allemagne, deux centres de numérisation rétrospective ont été fondés au sein de la « Niedersächsische Staats- und Universitätsbibliothek » à Göttingen <<http://gdz.sub.uni-goettingen.de/en/index.html>> et parallèlement un deuxième centre relié à la « Bayerische Staatsbibliothek » à Munich <<http://www.bsb-muenchen.de/mdz/>>. Le soutien approfondi des initiatives par le public est un témoin vivant de l'impact croissant des médias numériques. La DFG a formulé des directives générales concernant les thèmes et collections d'ouvrages se prêtant essentiellement à la numérisation (AG Inhalt, 1997).

Récemment, la « Fédération internationale des associations de bibliothécaires et d'institution » (IFLA) a publié des recommandations pour les projets de numérisation des collections appartenant à l'état dans les librairies et archives qui vont de la sélection par le processus de numérisation à la gestion et diffusion des collections tenant simultanément compte de l'organisation du flux de travail, des aspects légaux et du budget (IFLA, 2002). Plusieurs projets allemands concernent des parties spécifiques de l'œuvre de nos grands poètes et écrivains mais aussi des manuscrits illustrés et des encyclopédies ou lexiques, y compris le dictionnaire allemand des frères Grimm (Dörr, 2000; Schibel, 2000). Une numérisation parallèle en mode image et en mode texte est la bienvenue parce qu'elle

conserve l'aspect original et permet la recherche plein texte. Une saisie complète de texte n'est pourtant pas envisagée dans tous les cas parce qu'elle demande une qualité exceptionnelle des documents et des logiciels de reconnaissance de caractères (Altrichter, 2001).

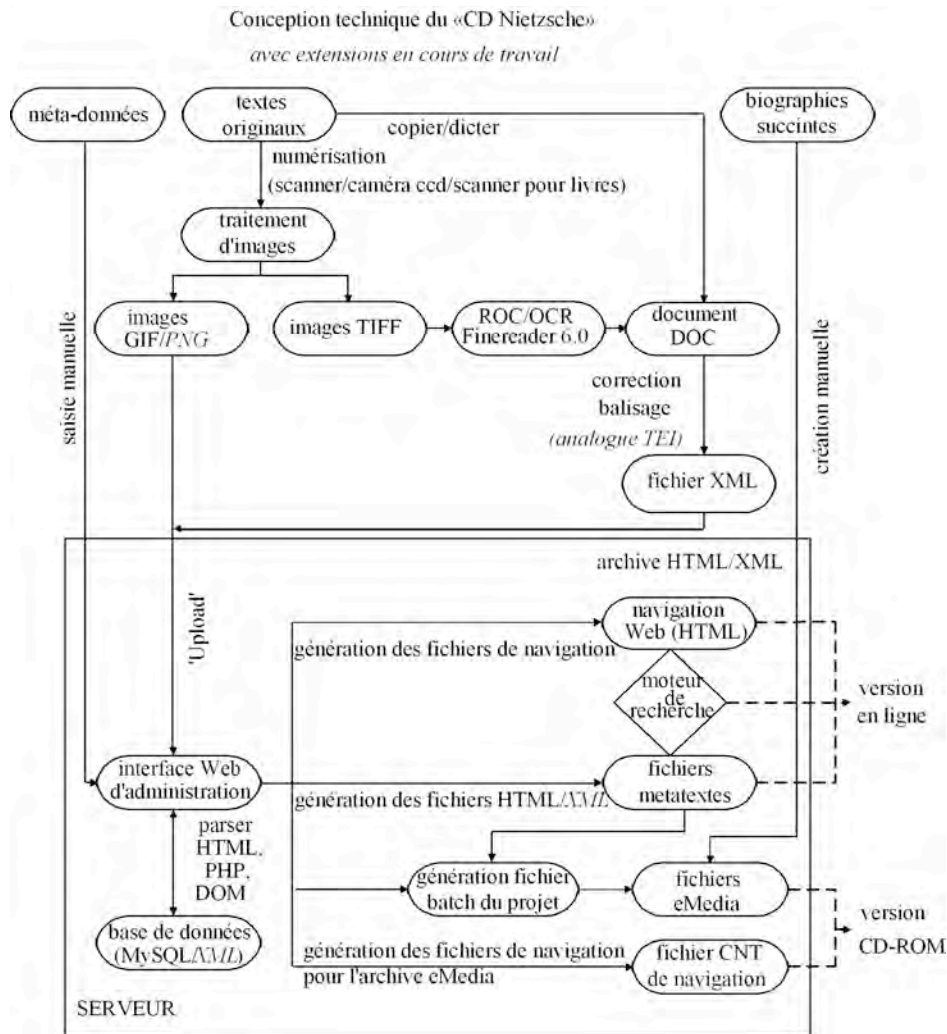
Citons aussi quelques initiatives internationales, tout d'abord le projet ALO (*Austrian Literature Online*) proposant une bibliothèque virtuelle de la littérature autrichienne (Egger & Mühlberger, 2000) soutenu par le projet européen META-e (Metadata Engine) <<http://meta-e.uibk.ac.at/>> (Miesenberger, 2002) et les actions de la bibliothèque du Congrès aux Etats-Unis portant le nom d'*American Memory* <<http://memory.loc.gov/ammem/amhome.html>>, de la Bibliothèque Nationale de France <<http://www.bnf.fr>> (Renoult, 1999) et de l'Electronic Text Center of the University of Virginia <<http://etext.lib.virginia.edu>>, fondé en 1992. Le projet américain « Journal Storage » (JSTOR) <<http://www.jstor.org>> est en quelque sorte un rediffuseur d'articles parus de deux à cinq ans auparavant qui effectue une numérisation rétrospective des collections.

D'autres projets de numérisation de l'héritage culturel du monde entier sont coordonnés par le répertoire des collections numériques sous l'égide de l'UNESCO et de l'IFLA. <<http://www.unesco.org/webworld/digicol/>>.

L'état actuel du projet de numérisation rétrospective de la réception de Nietzsche à Duisburg

Depuis l'an 2001, des chercheurs de différentes disciplines de l'université du Duisburg y compris l'informatique, les lettres et la linguistique, soutenus par le collège Friedrich Nietzsche à Weimar se sont rassemblés sous le nom de « CD Nietzsche » afin de créer des archives de la réception Nietzsche des années 1867 à 1945. Le plus grand travail a été effectué par les étudiants de la classe « sciences de communication et médias » sous l'égide de leurs professeurs. Depuis lors, près de 1800 pages de texte soigneusement choisies en fonction de leur accessibilité et de leur intérêt tant scientifique que culturel ont été répertoriées. Le choix des sources s'est orienté d'après les ouvrages de Richard Frank Krummel (1998) « Nietzsche und der deutsche Geist » et la toute récente « Weimarer Nietzsche Bibliographie » (2002).

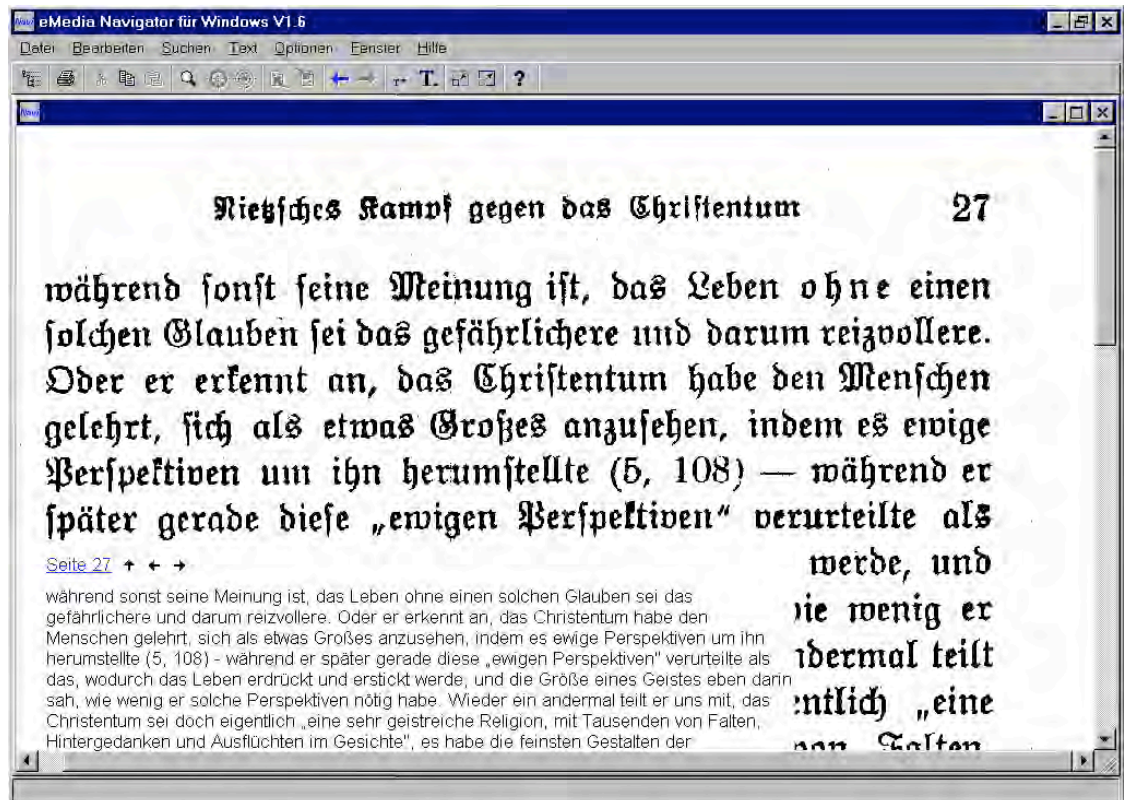
Nous numérisons les grandes lignes de la réception Nietzscheenne, des textes canoniques de caractère représentatif, soit d'une large répartition, soit d'une influence historique remarquable. Un premier effort a concerné les livres, extraits de journaux et textes canoniques de la réception religieuse et philosophique au cours de l'hiver 2001/2002, auxquels se sont ajoutées, un an plus tard, des œuvres d'auteurs socialistes, féministes ou antisémites.



Conception technique des archives de la réception Nietzscheenne

Un autre point d'intérêt était l'image de Nietzsche pendant la première guerre mondiale et les adeptes du culte de Nietzsche. Dans un futur proche, nous aimerions élucider l'implication des personnages influents antisémites Max Liebermann von Sonnenberg et Paul Förster au monde d'idées du philosophe Nietzscheen et leurs relations aux archives de Nietzsche. Les archives créées pendant les dernières années sont prometteuses et seront successivement complétées.

A l'heure actuelle, une centaine d'ouvrages et articles est numérisée, corrigée, encodée en méta-données et incorporée dans nos archives. Afin de conserver la mise en page originale, nous avons joint aux textes reconnus par ROCisation en couche additionnelle les formats d'images scannées et réduites à une taille convenable. Ainsi, nous disposons d'un accès simultané, d'une recherche plein texte et d'une vue sur la structure originale des documents.



Montage de texte numérisé et d'image fac-similé

Comme les logiciels de reconnaissance optique des caractères « Fraktur » ne procurent que des résultats modestes jusqu'à ce jour et d'autres initiatives comparables à la nôtre, y compris l'ambitieux projet européen META-e, n'ont pas abouti, nous avons élaboré un processus intégral comprenant la numérisation, la création de schémas mots-clés, la saisie des méta-données, correction et balisage, la transformation en document dans un format de documents standardisé de type «eXtensible Markup Language» XML/HTML, la mise en contexte et la gestion de l'archive (voir figure 1). Nous basant sur un échange d'expériences avec les auteurs de la bibliographie de Nietzsche (WNB, 2002), nous avons établi une classification hiérarchique de mots-clés selon nos thèmes. Nous avons utilisé un système de balisage de textes similaire au langage de balisage TEI lite (Burnard &

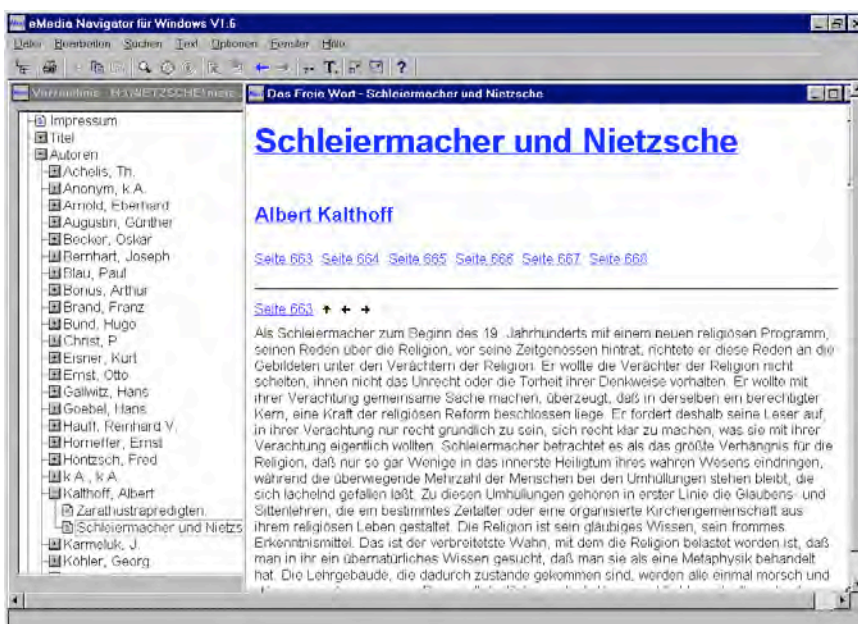
Sperberg-McQueen, 1995). L'exemple le plus connu de méta-données, le « Dublin Core Metadata Element Set » <<http://dublincore.org/>>, développé par une commission internationale et interdisciplinaire, constitué de quinze éléments de description des ressources en ligne, a également été pris en considération. Afin de valider le balisage manuel, nous avons élaboré une définition de document type (DTD).

DTD pour périodiques dans le projet CD-Nietzsche

```
<![IGNORE[<!ELEMENT artikel (literaturangaben?, seite*)>]]>
<![INCLUDE[<!ELEMENT artikel (seite*)>]]>
<!--Les données bibliographiques diffèrent pour livres et périodiques -->
<![IGNORE[
  <!ELEMENT literaturangaben (zeitung | buch)>
  <!ELEMENT zeitung (titel, undertitel, band, nr, seitenzahl*, erscheinungsdatum, verlag, herausgeber)>
  <!ELEMENT buch (titel, undertitel, band, auflage, verlag, herausgeber)>
  <!ELEMENT titel (#PCDATA)>
  <!ELEMENT undertitel (#PCDATA)>
  <!ELEMENT band (#PCDATA)>
  <!ELEMENT nr (#PCDATA)>
  <!ELEMENT seitenzahl EMPTY>
  <!ELEMENT erscheinungsdatum (#PCDATA)>
  <!ELEMENT verlag (#PCDATA)>
  <!ELEMENT herausgeber (#PCDATA)>
  <!ELEMENT auflage (#PCDATA)>
]]>
<!--Les pages contiennent l'information -->
<!ELEMENT seite (inhalt?, fussnote*)>
<!ELEMENT inhalt (#PCDATA | ueberschrift | fussnotenverweis | absatz | unleserlich)*>
<!ELEMENT ueberschrift (#PCDATA | absatz | fussnotenverweis | unleserlich)*>
<!ELEMENT fussnotenverweis EMPTY>
<!ELEMENT fussnote (#PCDATA | absatz | unleserlich)*>
<!ELEMENT unleserlich EMPTY>
<!ELEMENT absatz EMPTY>
<!-- Liste des attributs-->
<!ATTLIST artikel
artikel_id ID #REQUIRED
>
<!ATTLIST seite
seite_id ID #REQUIRED
label CDATA #REQUIRED
>
<!ATTLIST ueberschrift
ueberschrift_id ID #REQUIRED
tiefe (1 | 2 | 3 | 4 | 5 | 6) #REQUIRED
>
<!ATTLIST fussnotenverweis
fussnote_id ID #REQUIRED
>
<!ATTLIST fussnote
fussnote_id IDREF #REQUIRED
>
<![IGNORE[
<!ATTLIST seitenzahl
von CDATA #REQUIRED
bis CDATA #REQUIRED
>
```

]]>

Un scripte PHP basé sur la bibliothèque LibXML et les notions du Modèle Objet de Document DOM sert à saisir les données et les stocker dans une base de données MySQL. Afin de proposer parallèlement les archives en version CD et directement accessibles sur le Web, nous avons converti les données au format HTML. Après une étude comparative, nous avons retenu la Suite « eMedia Search » qui a été mise à notre disposition gratuitement par l'édition Heise. Elle enregistre les données au format HTML et dispose d'un moteur de recherche floue. En même temps que les fichiers HTML, un fichier de navigation est créé qui permet l'accès aux documents et la recherche par mots-clés.



Page HTML concernant un texte d'Albert Kalthoff

La transformation des documents du format XML au format HTML ne sera plus nécessaire à partir du moment où les navigateurs Web supportant l'XML et les éditeurs et bases de données XML natives seront plus répandus. C'est pourquoi nous utilisons d'ores et déjà l'éditeur XML-Spy et l'entreprise Infonyte nous équipera avec sa base de données XML dans un futur proche.

Quelques aspects techniques de la numérisation rétrospective

Nous mentionnons ici plusieurs problèmes qui nous ont forcés à développer des solutions nouvelles et importantes pour des activités de numérisation ressemblantes.

☒ La saisie est gênée par le fait que les documents à numériser sont dispersés dans toute l'Allemagne en forme de livre, périodique ou lettre. Les éditions utilisent une orthographe non standard et des polices « Fraktur » assez différentes. De plus, la qualité du papier et l'état de chaque document diffère grandement. Il a toujours fallu d'entraîner le logiciel ROC utilisé FineReader 6.0 sur chaque document, ce qui n'est guère efficace pour les textes d'une ou deux pages.

☒ La reconnaissance des majuscules en général composées de plusieurs arcs s'est avérée très difficile, et les ligatures *f* et *st*, des arcs ultra fins en *n* et *m* sont souvent mal re-

connus. Une correction manuelle peut être effectuée à l'aide d'un correcteur d'orthographe automatique mais demande souvent une analyse sémantique du texte.

die ihren Völkern Schicksal sind	„die ihren Völkern Schickfal sind“
immerhin die Weissagungen	„immerhin die Weissagungen“
Holz und steifbeinig in den Prunkjäten	„ftolz und steifbeinig in den Prunkjäten“ (Eisner 1891, 1661)
habe ich gehalten	„habe ich gehalten“
nur das Durchschnittsleben	„nur das Dnrchschnittsleben“ (Bonus 1899, 9)

Erreurs typiques dans la reconnaissance optique de caractères « Fraktur »

☒ Les logiciels de reconnaissance de caractères commercialisés permettent l'exportation des textes reconnus vers les formats courants DOC, RTF ou HTML, mais ne supportent pas la saisie des méta-données structurelles et descriptives en utilisant le langage de balisage XML ni ses dérivés comme TEI lite créée par la Text Encoding Initiative (TEI). Ainsi, les caractéristiques de la police « Fraktur », passages soulignés et espacés ne sont pas conservés. La pagination du texte respectant la structure de la police définie par la taille des caractères et l'espacement ne sont garantis que par une saisie image ou par l'enrichissement manuel du texte à l'aide d'un langage de balisage. La création des archives textes en images est donc une solution souvent choisie. Cependant, elle demande un traitement d'image élaboré afin de garantir une qualité égale et gêne fortement la recherche texte basée sur des schémas mots-clés et la comparaison des textes. L'alternative consiste à extraire et stocker les méta-données structurelles comprenant l'information sur

les auteurs, les titres et sous-titres, la composition du texte en chapitre, paragraphe, citations et annotations moyennant un balisage manuel ou semi-automatique.

☒ Un autre problème est lié au fait que la plupart des textes numérisés ont été écrits avant la réforme de l'orthographe en Allemagne en 1901/02 et emploient souvent l'orthographe phonétique. A ce propos, un moteur de recherche a été conçu permettant la recherche floue, phonétique et par troncation. Ceci s'est avéré nécessaire étant donné que les solutions commercialisées ne proposent qu'un support concernant les erreurs de frappe et l'emploi de caractères joker. Le logiciel est basé sur un algorithme qui déduit pour chaque nom, verbe et adjectif toute orthographe possible selon un système de règles linguistiques reportées dans la littérature ou dérivé en collaboration avec des spécialistes.

Les exemples suivants illustrant les effets d'une orthographe phonétique sont donnés dans l'ouvrage de Pilz (2003).

„das *grewliche thier* Luthers“ (Hauff 1939, 63)

„ihr *todten*, [...] ihr *rothen*“ (Brand 1898, 388)

„jeder Irrtum zu *Mißgriffen*“ (Frey 1887, 10)

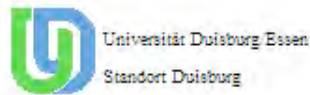
„das die *deutsche Litteratur*“ (Eisner 1891, 1506)

„sich am *herlichsten offenbart*“ (Eisner 1891, 1661)

L'auteur développe un jeu de règles afin de déduire toute forme d'orthographe possible des substantifs, verbes, adverbes et adjectifs et construit une base de données contenant les dérivées pour tous les mots importants des archives.

Le premier jeu concerne les voyelles à tréma et les diphtongues, le deuxième jeu déduit différentes règles et transformations phonétiques de lettres selon la coutume des époques en respectant les résultats de la mutation consonantique germanique et « haute-allemande ».

Par la suite, les déformations causées par une prononciation incorrecte ou influencées par l'origine de l'auteur et son lieu de naissance sont prises en considération.



"Auf den Spuren eines Philosophen".
Eine Quellensammlung zur Rezeption Friedrich Nietzsches in Deutschland von 1865-1945.



Tun Vorschau [Hilfe](#)
Suche: scharf unschärfer noch unschärfer

201. [TUN](#)

[in RELIGIÖSE GRUNDFRAGEN UM FRIEDRICH NIETZSCHE von THEODOR STEINBÜCHEL](#) [Treffergröße 1](#)

202. [THUN](#)

[in ZUM NIETZSCHE-PROBLEM von ARTHUR BONUS](#) [Treffergröße 2](#)

Interface de la recherche phonétique

La recherche est pilotée par une mesure de similitude et fournit les informations suivantes :

- Le mot trouvé
- Le titre du texte contenant le mot
- L'auteur du texte
- Similitude selon la distance de Levenshtein
- L'extrait du texte

Le moteur propose trois degrés de similitude :

1. La recherche précise « scharf » fournit uniquement les mots qui correspondent exactement à la demande.
2. La recherche floue du premier degré « unschärfer » sort tous les mots déduits du mot recherché selon les règles des transformations phonétiques ou historiques. Citons deux exemples : **TH** \square **T** („Thun“ \square „Tun“) ou **TZ** \square **Z** („Hertz“ \square „Herz“).
3. La recherche floue du deuxième degré « noch unschärfer » prend aussi en compte quelques règles spéciales et les erreurs de reconnaissance optique.

The screenshot shows a search engine interface. On the left, there is a logo for 'Universität Duisburg Essen Standort Duisburg' and a quote: '"Auf den Spuren eines Philosophen". Eine Quellensammlung zur Rezeption Friedrich Nietzsches in Deutschland von 1865-1945.' In the center, there is a stylized graphic of the word 'Nietzsche'. On the right, the search term 'Allzumenschliches' is entered in a search box. Below the search box, there are buttons for 'Suchen' and 'Vorschau', and radio buttons for search options: 'scharf', 'unschärfer', and 'noch unschärfer'. A 'Hilfe' link is visible in the top right corner.

38. ALLZUMENSCHLICHEN

in NIETZSCHE UND DIE FRAUEN von HEINRICH STERN **Treffergüte 2**

39. ALLZUMENSCHLICHEN

in NIETZSCHES STELLUNG ZU WEIB, LIEBE UND EHE von WALTER JESINGHAUS **Treffergüte 2**

40. ALLZUMENSCHLICHEN

in NIETZSCHES STELLUNG ZU WEIB, LIEBE UND EHE von WALTER JESINGHAUS **Treffergüte 2**

41. ALLZUMENSCHLICHEN

in NIETZSCHES STELLUNG ZU WEIB, LIEBE UND EHE von WALTER JESINGHAUS **Treffergüte 2**

42. ALLZUMENSCHLICHEN

in NIETZSCHES STELLUNG ZUR CHRISTLICHEN DEMUT von REINHARD V. HAUFF **Treffergüte 2**

43. ALLZUMENFCHLICHES

in NIETZSCHE 'S BEDEUTUNG FÜR UNSERE ZEIT von K.A. K.A. **Treffergüte 2**

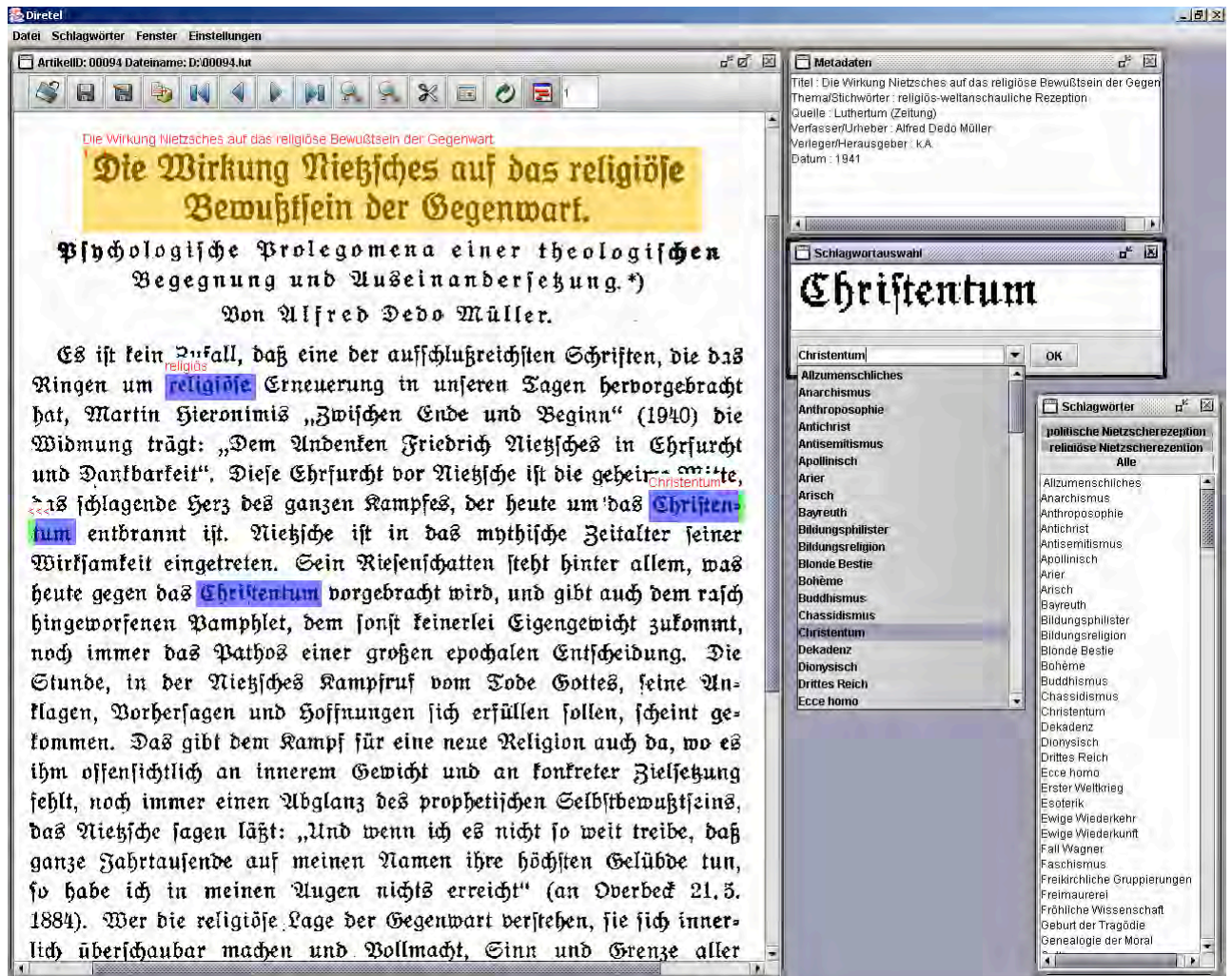
44. ALLZUMENSCHLICHEN

in FRIEDRICH NIETZSCHE UND DIE RELIGION von H. DRIESMANN **Treffergüte 2**

Extrait d'un résultat de la recherche floue du deuxième degré

Donnons un exemple :

Nous cherchons le mot « Allzumenschliches ». Le moteur de recherche fournit au degré trois le résultat de la figure 6 et détecte aussi l'erreur de numérisation « Allzumenfchliches ». La recherche floue incorporée dans le système eMedia Suite est très performante et dispose de plusieurs niveaux de similitude. La liste des mots sortis commence ainsi par « Allzumenschliches » suivi de « Allzumenschlicher, Allzumenschlichen, allzumenschlich, Menschliches, Übermenschlicher, unmenschlich, ... , fröhlich, schwächlich, schlicht » mais ne trouve ni les erreurs de numérisation ni les modifications phonétiques.



Interface de reconnaissance partielle du texte

☒ En outre, une interface a été conçue pour piloter la reconnaissance complète ou partielle du texte.

L'interface propose la navigation sur le texte d'un fichier numérisé et visualise des listes spécifiques des termes de réception gérées par des spécialistes (fenêtre « Schlagwörter ») qui sont réunies dans une liste globale « Alle ». Ceci permet la modification contrôlée des listes mots-clés et l'ajout d'autres champs de réception. Actuellement, nous disposons d'un index de mots-clés en structure arborescente concernant la réception religieuse et philosophique comprenant plus de 70 enregistrements qui est en permanence complété et élargi d'autres champs de réception. Nous pouvons largement profiter d'un échange d'expérience avec le directeur de l'édition de la Weimarer Nietzsche Bibliographie M. Erdmann von Wilamowitz-Moellendorff (WNB, 2002).

Un mot-clé qui apparaît dans le texte peut être sélectionné avec la souris, même en cas de césure. Le marquage visualisé par des blocs bleus dans le texte même et des blocs jaunes pour les entêtes de chapitres entraîne le stockage du mot-clé dans un fichier de type XML (voir exemple) avec quelques méta-données comprenant le titre du texte, la page et sa position dans un système de coordonnées. Ainsi, cette reconnaissance partielle du texte liée à la numérisation en images permet l'utilisation d'un moteur de recherche. La fenêtre « Metadaten » permettra l'édition des méta-données du document et leur stockage dans la base de données. Donnons un exemple du fichier XML des méta-données :

```
<?xml version="1.0" encoding="UTF-8"?>
<keywordposition xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="keywordposition.xsd" ID="KEYPOS1"
picpath="http://nietzsche.informatik.uni-duisburg.de/pictures/"
artID="00015">
....
<keyword ID="KEYP3" word="Christentum">

<marking ID="MARK3">
<coordinates ID="CO03" x="749" y="560" height="36" width="110"
path="http://nietzsche.informatik.uni-duisburg.de/pictures/00015_00001.png"
side="1" />
<coordinates ID="CO04" x="100" y="591" height="32" width="60"
path="http://nietzsche.informatik.uni-duisburg.de/pictures/00015_00001.png"
side="1" />
</marking>

<marking ID="MARK4">
<coordinates ID="CO05" x="279" y="655" height="31" width="139"
path="http://nietzsche.informatik.uni-duisburg.de/pictures/00015_00001.png"
side="1" />
</marking>
</keyword>
....
</keywordposition>
```

Figure 8 : Extrait du fichier keywordposition.xml

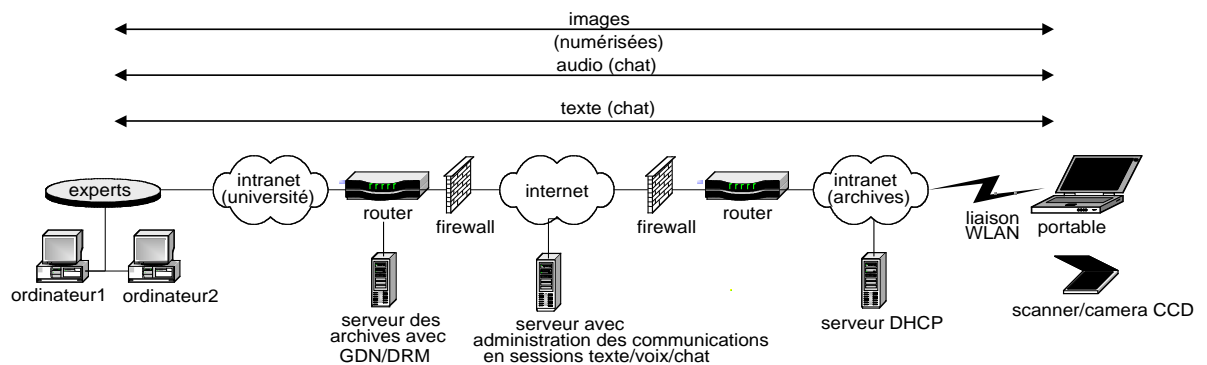
Recherche mobile et répartie de documents dans les archives et bibliothèques

Un des points faibles de notre accès à la numérisation est le rassemblement centralisé des textes à traiter dans le projet par la voie de courrier. La qualité des copies des sources s'est montrée très médiocre et inégale selon leur provenance et souvent non appropriée à un traitement ultérieur.

Afin de perfectionner le processus de sélection, de conversion et de correction des documents, nous avons récemment lancé un projet intitulé « Recherche mobile et répartie de

documents dans les archives et bibliothèques » <<http://www.edumedia.uni-duisburg.de/ecampus/>> qui est soutenu par le ministère BMBF (Biella et al., 2003). Il a pour but de développer un scénario comprenant des ordinateurs mobiles intégrés dans un réseau WLAN avec adresses client IP temporaires dans nos bibliothèques. Etudiants et chercheurs passent les documents au crible en les saisissant par caméra CCD ou scanner crayon. Les images sont retransmises par Internet au sein d'un comité de spécialistes à un lieu quelconque qui guide le chercheur et coordonne la sélection, le pré-traitement et la numérisation efficace des documents. Notons que le transfert est souvent limité à une communication textuelle à cause des restrictions dues à la configuration restrictive de la firewall (voir figure 7) et la gêne qui procure une communication audiovisuelle dans un endroit calme. Les hypothèses suivantes seront étudiées dans le contexte du projet de numérisation rétrospective :

Des ordinateurs portables munis d'une interface WLAN et intégrés dans le réseau de la bibliothèque par un serveur DHCP servent à saisir des documents à l'aide d'une caméra CCD. Le DHCP (Dynamique Host Configuration Protocol) est un service, qui sert à configurer des machines clientes à distance. La principale utilisation du DHCP est de donner aux gens qui les demandent des adresses IP (Internet protocol) dans un réseau.



Réseaux et communication entre lieu de saisie et experts

Les images transmises par Internet sont examinées par les experts et les étudiants sur place sont conseillés et dirigés vers les documents valables. Ainsi, les prêts de bibliothèques ou l'envoi de copies aux particuliers sont substitués par un processus efficace sans rupture de type de média. Les textes sont transformés en documents électroniques, enrichis en méta-données, stockés, catalogués et conservés de façon standardisée dans un environnement muni d'un serveur avec gestion des droits numériques. Si la saisie du texte intégral s'avère trop coûteuse, une saisie partielle peut être entretenue sur place et un

scanner pour livres peut être utilisé pour améliorer la qualité de l'image. Le scénario permet aussi bien un travail en groupe dans un milieu uniforme client/serveur qu'une recherche individuelle pour un diplôme d'un examen ou une publication d'un projet scientifique. Un premier essai, entrepris le 25 juin 2003 entre Duisburg et Weimar s'est avéré très prometteur. Le fruit de notre initiative pourrait être un traitement intégral du document optimisé réduisant les interventions manuelles et les erreurs humaines à un minimum.

Résumé

Nous avons rapporté les premiers résultats de projet de numérisation rétrospective de la réception Nietzschéenne des années 1865 à 1945. Un premier CD avec une centaine de sources est sorti accompagné d'une version en ligne. Une interface de saisie partielle a été développée, un scénario de recherche mobile et répartie de documents développé et testé. Ainsi, l'élargissement de thèmes de réception va de pair avec l'optimisation de la saisie, l'emploi des langages de balisage standardisés et de bases de données gérant des fichiers au standard XML.

Références

- AG Inhalt (1997). Bericht der Arbeitsgruppe Inhalt. Empfehlungen zur inhaltlichen Auswahl von Bibliotheksmaterialien für die retrospektive Digitalisierungen. Göttingen. <http://www.sub.uni-goettingen.de/ebene_2/vdf/empfehl.htm>
- Altrichter, H. (2001). Retrodigitalisierung in Deutschland. Versuch einer Zwischenbilanz. München, Januar 2001 (Kolloquium der Digitalisierungszentren). <<http://www.bsb-muenchen.de/mdz/forum/altrichter/index.htm>>□
- Bick, A. (2003). Ein Vorgehensmodell zur Retrodigitalisierung von Frakturtexten im Spiegel aktueller Retrodigitalisierungsprojekte. Duisburg. http://www.informatik.uni-duisburg.de/Studienprojekte/Nietzsche/Vorgehensmodell_Retrodigitalisierung_Nietzsche.pdf
- Biella, D., Dyllong, E., Kaiser, H., Luther, W., Mittmann, Th. (2002). Wege zur digitalen Erfassung der Nachwirkung Nietzsches in Deutschland von 1865 - 1945. Ein Arbeitsbericht zum Duisburger Retrodigitalisierungsprojekt. Colloque „Vom Umgang Nietzsches mit Büchern zum Umgang mit Nietzsches Büchern“, Weimar 23.09.-25.9.2002, à paraître dans les actes du colloque.
- Biella, D., Luther, W. (2003). Mobile verteilte Dokumentenrecherche in Bibliotheken und Archiven. Workshop 16: Notebook-Einsatz in der Aus- und Weiterbildung, GI-Tagung Frankfurt am Main 30.9.2003, GI-Edition Lecture Notes in Informatics (LNI).
- Burnard L. and Sperberg-McQueen, M. (1995). TEI Lite: An Introduction to Text Encoding for Interchange. <<http://www.tei-c.org/Lite/>>
- Dörr, M. (2000). Planung und Durchführung von Digitalisierungsprojekten. In Digitale Archive und Bibliotheken, hrsg. von Hartmut Weber und Gerald Maier. Stuttgart, 103-113.

- Egger, A., Mühlberger, G. (2000). ALO oder die virtuelle Bibliothek der österreichischen Literatur. Ein Arbeitsbericht. Bibliotheksdienst, Heft 6.□
- Krummel, R. F. (1998). Nietzsche und der deutsche Geist. Berlin, New York.
- IFLA (2002). GUIDELINES FOR DIGITIZATION PROJECTS for collections and holdings in the public domain, particularly those held by libraries and archives.
<<http://www.ifla.org/VII/s19/pubs/digit-guide.pdf>>
- Miesenberger, K. (2002). Bibliothèques électroniques et édition électronique pour les personnes privées d'accès à l'imprimé - État de l'art et perspectives d'avenir.
<<http://www.snv.jussieu.fr/inova/villette2002/act4b.htm>>□
- Mittmann, Th. (2001). Friedrich Nietzsche: Judengegner und Antisemitenfeind. Erfurt.
- Pilz, Th. (2003). Unschärfe Suche in Textdatenbanken mit nichtstandardisierter Rechtschreibung am Beispiel von Frakturtexten zur Nietzsche-Rezeption. Staatsexamensarbeit, Universität Duisburg-Essen.
- Renoult, D. (1999). Das Digitalisierungsprogramm der Bibliothèque nationale de France. Bibliothek, Forschung und Praxis 23 1, 69-75.
- Schibel, W. (2000). Aus der Bibliotheksreserve ins Internet. Die elektronische Publikation gedruckter Quellen der Frühen Neuzeit. MATEO Mannheimer Texte Online.
<<http://www.uni-mannheim.de/mateo/verlag/reports/schibellux.html>>
- Weimarer Nietzsche-Bibliographie (WNB) (2002) bearb. von Susanne Jung, Frank Simon-Ritz, Clemens Wahle, Erdmann v. Wilamowitz-Moellendorff, Wolfram Wojtecki. Bd. 1: Primärliteratur 1867-1998. Stuttgart, Weimar, 2000 - Bd. 2-5: Sekundärliteratur 1867-1998, 2002.□

Biographies

Daniel Biella, né le 06.08.1971, est, après avoir fait ses études de mathématiques et d'informatique, devenu un membre de l'institut d'informatique et des systèmes interactifs (IIS). Il s'intéresse principalement au graphisme, traitement d'image et de systèmes d'informations basés sur Internet.

Eva Dyllong, née le 19.12.1970, a fait des études de mathématiques, de philosophie et d'informatique. Elle prépare actuellement un doctorat à l'IIS et s'occupe particulièrement de l'imagerie numérique, de la robotique, de la géométrie assistée par ordinateur et du calcul scientifique.

Herbert Kaiser, né le 09.04.1941, est après ses études d'allemand et d'histoire, depuis 1984, professeur d'université de germanistique et d'histoire de la littérature à l'université de Duisburg. Il s'occupe principalement de l'enseignement de la littérature et de l'herméneutique littéraire.

Wolfram Luther, né le 12.01.1947, est actuellement professeur d'université en informatique à l'université de Duisburg-Essen. Après ses études en mathématiques et physique, il a enseigné les mathématiques à l'université technique d'Aix-la-Chapelle, à l'université de

Haute Alsace et à l'université d'Ulm. Il s'occupe des sujets d'informatique appliquée, du graphisme, traitement d'image et systèmes interactifs pour l'enseignement.

Thomas Mittmann, né le 17.09.1965, a fait des études d'histoire, de philosophie et de sciences politiques à l'université de Bochum où il prépare actuellement un doctorat sur la réception antisémite de Nietzsche. Depuis le semestre d'hiver 2001/2002, Mittmann est un des responsables du projet de numérisation rétrospective de la réception Nietzscheenne à Duisburg.

{biella,dyllong,luther}@informatik.uni-duisburg.de, he202ka@uni-duisburg.de,

Thomas.Mittmann@ruhr-uni-bochum.de