

# جست و جو در نسخه‌های خطی عربی با استفاده از ابر داده و گزارمان

■ عمر البانی <sup>۱</sup>
■ رشید بن سلیمان <sup>۲</sup>
■ نورالدین المخفی <sup>۳</sup>
■ بدرالدین آغوتان <sup>۴</sup>
■ نورالدین رئیس <sup>۵</sup>
■ ترجمه حمید کشاورز، عضو هیئت علمی گروه کتابداری و اطلاع‌رسانی دانشگاه سمنان

## چکیده

هدف غایی پژوهش حاضر بهبود شرایط نگهداری و دسترسی گسترده به میراث ارزشمند فرهنگی و ادبی در مجموعه‌های عمومی و خصوصی نسخ خطی است. در این مقاله روشی برای بازیابی نسخ خطی بر مبنای انتخاب ابر داده‌هایی خاص نسخ خطی عربی ارائه می‌شود. این ابر داده‌ها به مثابه فهرست اولیه کلیدواژه‌های دسترسی به این نسخ خطی به کار گرفته شده‌اند. در ادامه به منظور ایجاد زمینه لازم برای بهبود جست‌وجوی اطلاعات، این فهرست با گزارمان‌هایی به صورت صفحات دیجیتالی تکمیل می‌گردد.

**کلیدواژه‌ها:** نسخ خطی، دیجیتالی‌سازی، جفت‌سازی متن/تصویر، ابر داده، گزارمان، آوانویسی، زبان نشانه‌گذاری گسترش‌پذیر، اصول کدگذاری متن.<sup>۷</sup>

## ۱. مقدمه

و ابرمتن<sup>۱۰</sup>، هنوز کمیابند. دلیل این امر شاید ماهیت پیچیده این کار یا کمبود ابزارهای حرفه‌ای برای پردازش نسخ خطی قدیمی و جدید باشد. (استفان<sup>۱۱</sup>، تیبری<sup>۱۲</sup>، لارنت<sup>۱۳</sup>، ۲۰۰۳). در بخش زیر آخرین پروژه‌های مربوط به دیجیتالی‌سازی نسخ خطی ارائه می‌گردد:

پروژه دسترسی بهتر به نسخ خطی و جست‌وجو در تصاویر<sup>۱۴</sup> (کالابرتو<sup>۱۵</sup>، بوزی<sup>۱۶</sup>، پینون<sup>۱۷</sup>، ۱۹۹۹) به یک سیستم ابرسازمانه‌ای انجامید که به مورخان و به‌ویژه نسخه‌شناسان و واژه‌شناسان امکان می‌دهد به مطالعه و آوانویسی نسخ خطی پرداخته و برای آنها گزارمان تهیه کنند. به‌علاوه امکان جست‌وجو بین کلمات آوانویسی شده و یافتن تصویر به در شکل عددی نسخ خطی فراهم آمده است.

پروژه دسترسی دیجیتالی به کتاب‌های دوره رنسانس<sup>۱۸</sup> (دبور<sup>۱۹</sup>، ۲۰۰۰) به دنبال رفع نیازهای مربوط به کاربرد کتابخانه‌های مجازی و کتاب‌های الکترونیکی و هدف آن توسعه ابزارهایی برای دیجیتالی‌سازی و دسترسی به گزیده‌ای از کتاب‌های قرن ۱۶ میلادی بود که منجر به ایجاد محیطی برای مطالعات مشترک بر روی مجموعه‌های دیجیتالی شد. هرچند این محیط نسخ خطی اصلی را ارائه نمی‌داد، طی این پروژه برخی جنبه‌های فنی مهم

خواندن مدارک آرشیوهای عربی مانند دیگر مدارک آرشیوی سرتاسر جهان با دشواری همراه است. دسترسی به این نسخ خطی مهم بسیار محدود است زیرا بسیاری از این مدارک در شرایط آسیب‌پذیری بوده و نیز نسخه‌برداری دستی از آنها بسیار زمان‌بر و طاقت‌فرسا است.

برای حل این مشکل سیستم‌هایی مورد نیازند تا بازیابی مدارک را از طریق محتوای شان امکان‌پذیر سازند. به این منظور ضروری است که از گزارمان‌ها [یادداشت‌های توضیحی] ابر داده‌هایی برای تصاویر این مدارک کمک گرفته شود. گزارمان‌های مدارک آرشیوی می‌توانند شکلی هندسی<sup>۸</sup> داشته باشند که نیازمند جایی در تصاویر آن مدارک است که به صورت مستطیل‌هایی همراه با گزارمان‌های متنی نمایش داده می‌شوند. البته می‌توان گزارمان‌های متنی را به شکل هندسی پیوند داد تا انتخاب تصویر مناسب به صورت خودکار امکان‌پذیر گردد.

## ۲. کارهای مشابه

نسخه‌های الکترونیکی آثار تاریخی به‌رغم تکامل فناوری‌های چندرسانه‌ای و فرصت‌های پدیدآمده توسط زبان‌های ساختارمند<sup>۹</sup>

جدول ۱. ابرداده

شماره ابرداده	ابرداده پذیرفته شده
۱	نویسنده نسخه نویس نام مالک نام مجموعه دار
۲	عنوان عنوان نسخه خطی عنوان فصل‌ها عنوان فصل‌های فرعی
۳	دوره مورد مطالعه بیت المقدس اسلامی قبل از اسلام قرن‌های ۴-۵ هجری قرن‌های ۶-۷ هجری قرون میانه اسلامی ۷-۹ هجری
۴	گروه نسخه خطی عربی-اسلامی عربی-مسیحی
۵	مواد مدرک نوع حمایت تاریخ مدرک
۶	نوع نسخه خطی قرآن دیگر متون دینی دانشمندان پزشکان ادبی فقه اسلامی فلسفی تاریخی سایر
۷	نگارش سبک نگارش
۸	ظاهر متون تذهیب تصاویر لوح آغازین

**پروژه دسترسی الکترونیک به نسخ  
خطی دوره قرون وسطی، پروژه‌های  
سه ساله در توسعه رهنمودهایی  
برای کدگذاری، ذخیره‌سازی و  
توصیف نسخ خطی دوران قرون  
وسطی و رنسانس به شکل  
الکترونیکی است**

مانند فرمت و فشرده‌سازی تصاویر مورد ملاحظه قرار گرفتند. پروژه دسترسی الکترونیک به نسخ خطی دوره قرون وسطی،<sup>۲۰</sup> پروژه‌های<sup>۲۱</sup> سه ساله در توسعه رهنمودهایی برای کدگذاری، ذخیره‌سازی و توصیف نسخ خطی دوران قرون وسطی و رنسانس به شکل الکترونیکی است. اهداف پروژه عبارتند از: (۱) شناسایی نوع اطلاعات موجود در رکوردهای کوچک فهرست‌نویسی و تعیین چگونگی قابلیت کدگذاری این اطلاعات برای ذخیره‌سازی در پایگاه‌های اطلاعاتی رابطه‌ای؛<sup>۲۲</sup> (۲) شناسایی و طبقه‌بندی اطلاعات اضافی موجود در یک رکورد کامل فهرست‌نویسی و تعیین چگونگی قابلیت کدگذاری این اطلاعات؛<sup>۲۳</sup> (۳) آزمایش و بهبود رهنمودهایی برای کدگذاری با ایجاد یک پایگاه اطلاعاتی از رکوردهای الکترونیکی که مطابق با این رهنمودها کدگذاری شده باشند و دسترسی برخط به این پایگاه اطلاعاتی را میسر سازند.

**عنصر ریشه‌ای در  
فایل زبان نشانه‌گذاری  
گسترش پذیر،  
«توصیف نسخه خطی»  
نامیده شده و عنصری  
است که قاعدتاً در  
متن یا در عنوان چنین  
فایلی مطابق با تی.ای.  
آی می‌آید**

**۴. نمودار طرح ابرداده‌ای  
عنصر ریشه‌ای**

عنصر ریشه‌ای در فایل زبان نشانه‌گذاری گسترش‌پذیر، «توصیف نسخه خطی»<sup>۲۴</sup> نامیده شده و عنصری است که قاعدتاً در متن یا در عنوان چنین فایلی مطابق با تی.ای. آی می‌آید.<sup>۲۵</sup> در حالت اول که مدرک در حال کدگذاری مجموعه‌ای از توصیفات نسخه خطی است محل قرار گرفتن عنصر توصیف نسخه خطی جاهایی است که پاراگراف ظاهر می‌شود. در حالت دوم که توصیف بخشی از ابرداده است که برای نمایش دیجیتالی برخی نسخه‌های خطی اصیل به کار برده می‌شود، توصیف نسخه خطی عناصر زیر را در بر دارد که تنها اولی ضروری است. این عنصر شامل هفت عنصر دیگر است که هر یک پیچیدگی خاص خود را دارد:

**۳. ابرداده**

نمایه‌های کتابداران برای توصیف مدارک موجود، تاریخی طولانی دارد. رکوردهای ابرداده‌ای حاوی اطلاعاتی هستند که برای توصیف اطلاعات دیگری همچون محتوای کتاب‌ها به کار برده می‌شوند؛ به همین خاطر به آنها ابرداده می‌گویند. ما می‌توانیم برخی از اطلاعات ضروری مدارک را به شیوه‌ای روشن کدگذاری کنیم شامل: عنوان، نویسنده، تاریخ انتشار، کلیدواژه‌ها و مانند آن. جدول ۱ چند دسته از ابرداده‌های به کار رفته در پروژه حاضر را به نمایش گذاشته است که قبلاً در گزارش (هالا<sup>۲۳</sup>، ۲۰۰۴) آورده شده است:

### عنصر تاریخ

عناصر گروهی برای توصیف تاریخ کامل یک نسخه خطی یا بخشی از آن.

عنصر تاریخ	
نام مالک	
نام مجموعه‌دار	
انتقال	
مهر	
مبدا	
منشاء	
فراهم‌آوری	
سایر	

شکل ۴. نمودار عنصر تاریخ

### عنصر محتوای نسخه خطی

این عنصر برای توصیف محتوای فکری یک نسخه خطی یا بخشی از آن به کار می‌رود. این عنصر می‌تواند شامل گروهی از پاراگراف‌های نوشتاری غیررسمی یا گروهی از عناصر ساختاریافته‌تر باشد که هر یک توصیف جزئی‌تری از یک عنصر موجود در نسخه خطی را فراهم می‌آورد.

عنصر محتوا	
نویسنده	
نسخه‌نویس	
عنوان	
صفحه‌شمار	
خط	
خلاصه	
موضوع	
انجامه	

شکل ۵. نمودار عنصر محتوای نسخه خطی

### عنصر ساختار منطقی

این عنصر ساختار منطقی مدرک، محتوا، بخش‌ها، فصل‌ها و غیره را توصیف می‌کند:

### عنصر توصیف نسخه خطی

شناسگر توصیف نسخه خطی
توصیف فیزیکی
تاریخ
محتوا
ساختار منطقی
اطلاعات مدیریتی
سایر

شکل ۱. نمودار عنصر توصیف نسخه خطی

### عنصر شناساگر نسخه خطی

این عنصر شامل همه عناصری است که به شناسایی نسخه خطی یا بخشی از آن کمک می‌کند:

### عنصر شناساگر نسخه خطی

مسئولیت
عنوان
مکان اصلی
مکان
تاریخ اصلی
طول متن
گروه نسخه اصلی
موضوع نسخه خطی

شکل ۲. نمودار عنصر شناساگر توصیف نسخه خطی

### عنصر توصیف فیزیکی

این عنصر تعداد زیادی از جنبه‌های کارآمد در توصیف یک نسخه خطی را پوشش می‌دهد. این جنبه‌ها شامل: شکل، نوع حمایت، طول، ساختار، دسته‌های کاغذی نسخه خطی، جنبه‌های نوشتاری و مانند آن است.

### عنصر توصیف فیزیکی

شکل
حمایت
طول
صفحه‌شمار
برگ‌ها
صفحه‌رایی
نگارش
تذهیب
صحافی
شرایط نگهداری
افزوددها

شکل ۳. نمودار عنصر توصیف فیزیکی

### ۵. شیوه پیشنهادی برای دسترسی به نسخه خطی

۱- ۵. چارچوبی برای گزارمان‌های مربوط به تصاویر برای دسترسی محتوایی به تصویری از یک مدرک باید به تصویر گزارمان‌هایی افزود (برتراند<sup>۲۷</sup>، ایوان<sup>۲۸</sup>، ۲۰۰۳). ما پیشنهاد می‌کنیم برای بازیابی مدارک آرشیوها که با گزارمان‌های متنی و هندسی در سطحی مشابه برخورد می‌کند چارچوبی ساخته شود؛ این چارچوب می‌تواند بین گزارمان‌های متنی و هندسی رابطه برقرار کند تا نشان دهد این اطلاعات متنی در چه قسمت خاصی از مدرک قرار دارند (شکل ۹). از آنجا که اطلاعات متنی زیادی در یک قسمت خاص از مدرک قابل یافتن است، باید حتی‌الامکان پیوندهای ضروری بین گزارمان‌های متنی و هندسی برقرار شود. یا اینکه یک گزارمان متنی بتواند به قسمت‌های مختلف در صفحات مختلف یک مدرک برای تصاویر متفاوتی در نظر گرفته شود و به آنها پیوند بخورد.

این چارچوب همچنین قادر است فایل‌هایی به گزارمان‌ها ضمیمه کند که برای ایجاد ابرداده‌های مربوط به تصاویر نسخه خطی قابل استفاده‌اند. ابرداده یا گزارمان ایجاد شده می‌تواند در فرمت زبان نشانه‌گذاری گسترش‌پذیر ذخیره شده و به عنوان اطلاعات ورودی برای بخش ۵ - ۲ استفاده شود. ما می‌توانیم این چارچوب را به عنوان برنامه‌ای برای دسترسی و کاربرد اطاعات موجود در نسخ خطی و نیز تصاویر آنها مورد توجه قرار دهیم.



شکل ۹. رابط کاربری برای گزارمان نویسی نسخه خطی

### ۵ - ۲. ابزاری برای کار بر روی مدارک ساختاریافته

#### کلیات سیستم

هدف پروژه «سیستمی برای مدارک در بستر نحوی زبان نشانه‌گذاری گسترش‌پذیر»<sup>۲۹</sup>، ایجاد سیستمی با پیروی از «گواهینامه عمومی»<sup>۳۰</sup> به صورت منبع باز است. این سیستم یک موتور جست‌وجو را به دست می‌دهد که مستقیماً آدرس‌های اینترنتی سپردنگاه‌های<sup>۳۱</sup> نسخ خطی را نمایه می‌کند. به علاوه اجازه تعریف یک یا چند مدل نمایه‌سازی (سپردنگاه‌های مدارک) را با نظم ربطی<sup>۳۲</sup> و بر اساس سطح خود فراهم آورد تا مدارک به نمایش درآیند بدون اینکه نسخه‌های اصلی تعدیل شوند. این سیستم یک سرولت جاوا<sup>۳۳</sup> است که گسترشی است از سیستم کوکون<sup>۳۴</sup> با استفاده از موتور

#### عنصر ساختار منطقی

صفحه عنوان
فهرست مندرجات
نمایه
مقدمه
بخش
فصل
انجامه
نویسنده همکار

شکل ۶. نمودار عنصر ساختار منطقی

#### عنصر اطلاعات مدیریتی

اطلاعات بسیاری در ارتباط با نگهداری و مدیریت یک نسخه خطی با استفاده از نوشته‌هایی ساده با کاربرد عنصر استاندارد P قابل ذخیره‌سازی هستند:

#### عنصر اطلاعات مدیریتی

P
تاریخ ثبت
موجودیت
تاریخ نگهداری
نشانه‌ها

شکل ۷. نمودار عنصر اطلاعات مدیریتی

#### عنصر سایر

این عنصر دیگر اطلاعات مرتبط به نسخه خطی به ویژه اطلاعات مدیریتی مربوط به مکان کنونی آن، مواد اضافه همراه با آن و غیره را در بردارد:

#### عنصر سایر

P
نسخه دیجیتال
نسخه میکروفیلمی
نسخه عکسی
نسخه انتشار

شکل ۸. نمودار برای عنصر سایر



**ناویمیج‌ها محیط‌هایی  
انعطاف‌پذیر، باز  
و استاندارد برای  
سازماندهی، پردازش  
و انتشار تصاویر فراهم  
می‌کنند، در عین حال  
کارکردهای بازیابی،  
مرور، و جست‌وجو را  
نیز دارا هستند**

جست‌وجوی لاسن<sup>۳۵</sup> و یک پردازشگر با نام کاربرگه گسترش‌پذیر تغییرات زبانی.<sup>۳۶</sup> بنابراین کاربرد این چارچوب فراتر رفتن از سیستم کاربرگه گسترش‌پذیر تغییرات زبانی است؛ یعنی مجموعه‌ای از صفحات با استفاده از فناوری‌های صفحات سرور گسترش‌پذیر<sup>۳۷</sup> و کاربرگه گسترش‌پذیر تغییرات زبانی که با کارکردهای متفاوت به کاربران ارائه می‌شود؛ از جمله جست‌وجو، فهرستی از اصطلاحات، ارائه نتایج یا مدارک و مانند آن.



شکل ۱۲. نمایش ابردادهای نسخه خطی



شکل ۱۰. مرور بر اساس موضوع



شکل ۱۱. فهرست عناوین نسخه های خطی

### ۵ - ۳. ناویمیج(ها)<sup>۳۸</sup>

ناویمیج‌ها (۱۱) محیط‌هایی انعطاف‌پذیر، باز و استاندارد برای سازماندهی، پردازش و انتشار تصاویر فراهم می‌کنند، در عین حال کارکردهای بازیابی، مرور، و جست‌وجو را نیز دارا هستند. هر ناویمیج از سه جزء سازنده<sup>۳۹</sup> برای دسترسی به سه هدف زیر بهره می‌گیرد: سازماندهی و پردازش تصویر، جست‌وجو در توصیفات تصویر و مرور و گردش در تصاویر در یک محیط توزیع شده. ناویمیج‌ها به‌طور بالقوه به سازماندهی و پردازش تصاویر می‌پردازند؛ هنگامی که تعریفی از نوع تصاویر به آنها داده شود (جزء گردآورنده)، امکان پیوند بین مجموعه تصاویر به فناوری‌های جست‌وجو (جزء مدرک) را فراهم آورده و محیطی برای مشاهده تک تک تصاویر هنگام مرور مجموعه‌ای از تصاویر (جزء نمایش) به‌وجود می‌آورند. این سه جزء سازنده بر اساس فناوری‌های منبع باز برای ایجاد انعطاف‌پذیری بیشتر در محیط‌های مختلف و در نتیجه سازگاری برای پروژه‌های آرشویی متفاوت ساخته شده‌اند.

17. PINON
18. Digital access to Books of Renaissance: DEBORA
19. DEBORA
20. Electronic Access to Medieval Manuscripts: EAMMS
21. <http://www.hmml.org/eamms/index.html>
22. Relational database
23. Hala
24. msDescription
25. Text Encoding Initiative
26. <http://www.tei-c.org/index.xml>
27. Bertrand
28. Ivan

29. System for Documents in XML: SDX (<http://adnx.org/>)
30. General Public License
31. Repositories
32. Relevance ordering
33. Servlet Java
34. Cocoon
35. Lucene
36. eXtensible Stylesheet Language Transformations: XSLT (<http://www.w3.org/TR/xslt>)
37. eXtensible Server Pages (<http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages/>)
38. Navimages
39. Module



شکل ۱۳. نمایش کلی نسخه خطی



شکل ۱۴. کاربران به راحتی می توانند بر روی نسخه خطی تمرکز کرده و بررسی های کلی و جزئی انجام دهند.

در این مقاله تلاش کردیم چارچوبی برای بهبود دسترسی به نسخ خطی عربی معرفی کنیم. برای دسترسی سریع، بر داده ها و گزارمانها ضروری هستند. به علاوه نشان دادیم که گزارمانهای مدارک آرشیوی می توانند هندسی باشند

مآخذ

[1] Stéphane Nicolas, Thierry Paquet, Laurent Heutte, 2003. "Digitizing Cultural Heritage Manuscripts": the Bovary Project. in ACM Symposium on Document Engineering, ACM Doc Eng 2003, Grenoble, France, pp. 55-57.

[2] CALABRETTO, Sylvie; BOZZI, Andrea; PINON, Jean-Marie, décembre 1999. "Numérisation des manuscrits médiévaux": le projet européen BAMBI, in: Actes du colloque Vers une nouvelle érudition: numérisation et recherche en histoire du livre, Rencontres Jacques Cartier, Lyon.

[3] DEBORA: projet européen n°. LB 5608 A. Coordinateur R. Bouché, juin 2000. 179 pages.

[4] <http://www.hmml.org/eamms/index.html>

[5] Hala kaileh, 2004. "L'accès à distance aux manuscrits arabes numérisés en mode image". Thèse présentée devant l'université Lumière Lyon II.

[6] <http://www.tei-c.org/index.xml>

[7] Bertrand Coüasnon, Ivan Leplumey. "A Generic Recognition System for Making Archives Documents accessible to Public". Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003) 0-7695-1960-1/03 \$17.00 © 2003 IEEE.

[8] <http://adnx.org/>

[9] <http://www.w3.org/TR/xslt>

[10] <http://java.apache.org/cocoon/xsp/WD-xsp.html>

[11] <http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages/>

ع. نتیجه گیری

در این مقاله تلاش کردیم چارچوبی برای بهبود دسترسی به نسخ خطی عربی معرفی کنیم. برای دسترسی سریع، بر داده ها و گزارمانها ضروری هستند. به علاوه نشان دادیم که گزارمانهای مدارک آرشیوی می توانند هندسی باشند.

پی نوشتها

1. Omar El bannay
2. Rachid Benslimane
3. Nouredine El makhfi
4. Badraddine Aghotane
5. Nouredine Rais
6. XML
7. Text Encoding Initiative: TEI
8. Geometric
9. Structured languages
10. Hypertext
11. Stéphane
12. Thierry
13. Laurent
14. Better Access to Manuscripts and Browsing of Images: BAMBI
15. CALABRETTO
16. BOZZI