



University Libraries



Al-Ādāb Magazine Archives: Digitization, Preservation and Access

BASMA CHEBANI (EMAIL: bc01@aub.edu.lb)

HEAD OF CATALOGUING AND METADATA SERVICES, UNIVERSITY LIBRARIES

ELIE KAHALE (EMAIL: ekahale@aub.edu.lb)

HEAD OF DIGITAL INITIATIVES AND IMAGING DEPARTMENT, UNIVERSITY LIBRARIES

AUB University Libraries



AUB University Libraries are leading different **digitization initiatives** in attempt to preserve National and AUB cultural heritage, in addition to disseminate information and promote knowledge by allowing access to AUB community, to scholars, to researchers and to the largest possible audience.

Digitization Initiatives - ACO



- Arabic Collections Online (ACO): is a publicly available digital library of public domain Arabic language content the project is managed by NYU.
- ACO contributing partners are New York University, Princeton, Cornell, Columbia, American University of Beirut and American University of Cairo.

Digitization Initiatives – Oral History



- **Palestinian Oral History Archive:** is an archival collection that contains more than 1,000 hours of testimonies with first generation Palestinians and other Palestinian communities in Lebanon.
- The project will digitize, index, catalog, preserve, and provide access to the material through the creation of a digital platform. It aims to expand and include additional oral history collections documenting varied aspects of the Palestinian experience in Lebanon and the region. The project is being completed by the AUB Libraries in partnership with the Issam Fares Institute for Public Policy and International Affairs at AUB, the Nakba Archive and the Arab Resource Center for Popular Arts (AL-JANA).

Digitization Initiatives – Nahda Journals



- Indexing the Nahda journals with Brill: AUB UL signed an agreement with Brill Publishers to start a joined project between the UL and Brill to index the Arabic Nahda journals using Index Islamicus platform and indexing techniques.
- The project proposes indexing the journal content in Arabic and English language, developing an **ontology** of Nahda term and concepts, and building authority lists of proper names of people, places and organizations.

Digitization Initiatives – AL ADAB



DIGITIZING AL-ĀDĀB MAGAZINE ARCHIVES

Agreement between Al-Ādāb and AUB University Libraries



- Al-Ādāb Magazine (مجلة الآداب) and AUB Libraries **agreed** in 2014 to **digitize, OCR, catalog** and **index** the entire magazine.
- Al-Ādāb Magazine will **publish** the digitized content through their own Content Management System (CMS)

Agreement between Al-Ādāb and AUB University Libraries



- AUB will keep a **preservation** copy of the entire magazine issues.
- University Libraries will provide the AUB community with access to the digitized content through a **full text search web interface**.

Why Al_Ādāb Magazine?



- Literary and cultural Journal **established in 1953** by Dr. Souhail Idris
- Focused on **movements in literature and culture in the Arab world**
- Included files in **political thought, poetry, novel, short stories, movies criticism, theater, and general culture.**

Why Al_Ādāb Magazine?



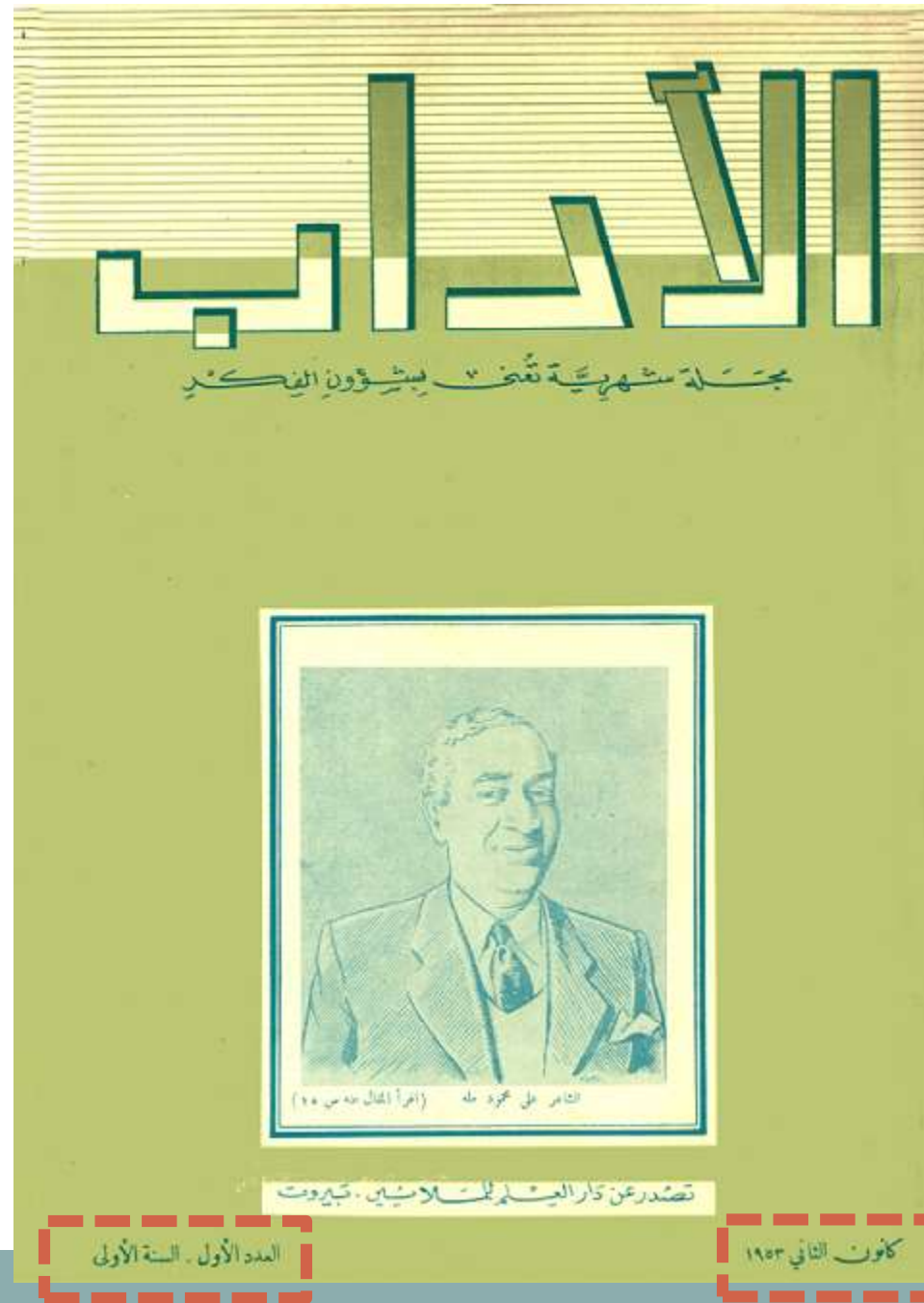
- Considered as **Arab cultural platform**, despite the Arab States censorship
- Al_Ādāb Magazine stopped publishing in print in 2012. Accumulating an archive of 60 volumes over 60 years.
- In October, 2015 Al_Ādāb Magazine started re-publishing only online.

الأداب : مجلة شهرية تعنى بشؤون الفكر



صدرت «الأداب» في عام ١٩٥٣م عن دار العلم للملايين، بعد عودة صاحبها ورئيس تحريرها الأول سهيل إدريس إلى بيروت، قادماً من باريس، حيث تخرج في جامعتها . أنشأها سهيل إدريس في البداية مع الناشرين منير بعلبكي وبهيج عثمان ثم استقلّ بها حيث أصبحت تصدر عن دار الأداب. بعد ذلك سرعان ما عرفت انتشاراً واسعاً في العالم العربي لأسباب مختلفة منها توجهها الفكري ك **مجلة عروبية تقدمية** مواكبة للمد القومي العربي الذي تصاعد في تلك المرحلة، ومنها أيضاً انفتاحها على التيارات الفكرية الجديدة مثل «الالتزام» و«الواقعية» والفكر الوجودي الذي أقبل عليه كثيرون لأنه كان، عكس الماركسية، يوفق بين «الحرية» و «الالتزام». وعلى مدى سنوات طويلة كانت «الأداب» هي المسوّق الأول لهذا الفكر عربياً، حيث صدر عن دارها «دار الأداب» ما لا يحصى من الكتب الوجودية الفرنسية وفي طليعتها كتب سارتر وألبير كامو وسيمون دو بوفوار. وقد صدر قسم كبير من هذه الكتب بترجمة سهيل إدريس نفسه. توقفت عن الصدور مع سنتها الستين في نهاية عام 2012 مع آخر عدد لها في خريف 2012 .

Cover of the **first issue** from the **first publication year**
1953



Title of the Magazine



- Title varied during the Al-Ādāb Magazine life time.

○ الآداب : مجلة شهرية تعنى بشؤون الفكر

○ الآداب : مجلة ثقافية عربية

○ الآداب

- Issued on monthly basis from 1953-1980
- Issued 5 times a year from 1980-2011.
- Issued 4 times a year in 2012
- Stopped by the end of Autumn 2012.

Al-Ādāb Magazine Archives



DIGITIZATION AND OCR

Why Digitization?



- Digitization for **preservation** (Digital Preservation)
 - **Authenticity**: Maintain trustworthy representation of original document.
 - **Integrity**: Ensure data is recorded/retrieved exactly as intended.
 - **Discovery** and **identification**: Ability to locate item.
 - Continuous ability to **use** and **access** a digital object.

Digital Preservation



- Initial steps for digital Preservation:
 - All Items were put in **bags** using **BagIt** and backed up on multiple copies on **SAN** storage and **Tapes** (in bit stream)
 - Library of Congress **BagIt** is a hierarchical file packaging format designed to support disk-based storage and network transfer of arbitrary digital content. “Wikipedia”

Digital Preservation



- Next steps: Ingest the bags into **Open Archival Information System**.
 - **Archivematica** is an open source digital preservation system which uses BagIt to create OAIS Archival Information Packages (**AIP**) “Wikipedia”
 - In Archivematica 1.4 and higher, fields in the bag-info.txt file are indexed as source metadata in the Archivematica **METS** file, making their contents **searchable** in the Archival storage tab after a bag transfer has been processed and stored. “Archivematica”

Why Digitization?



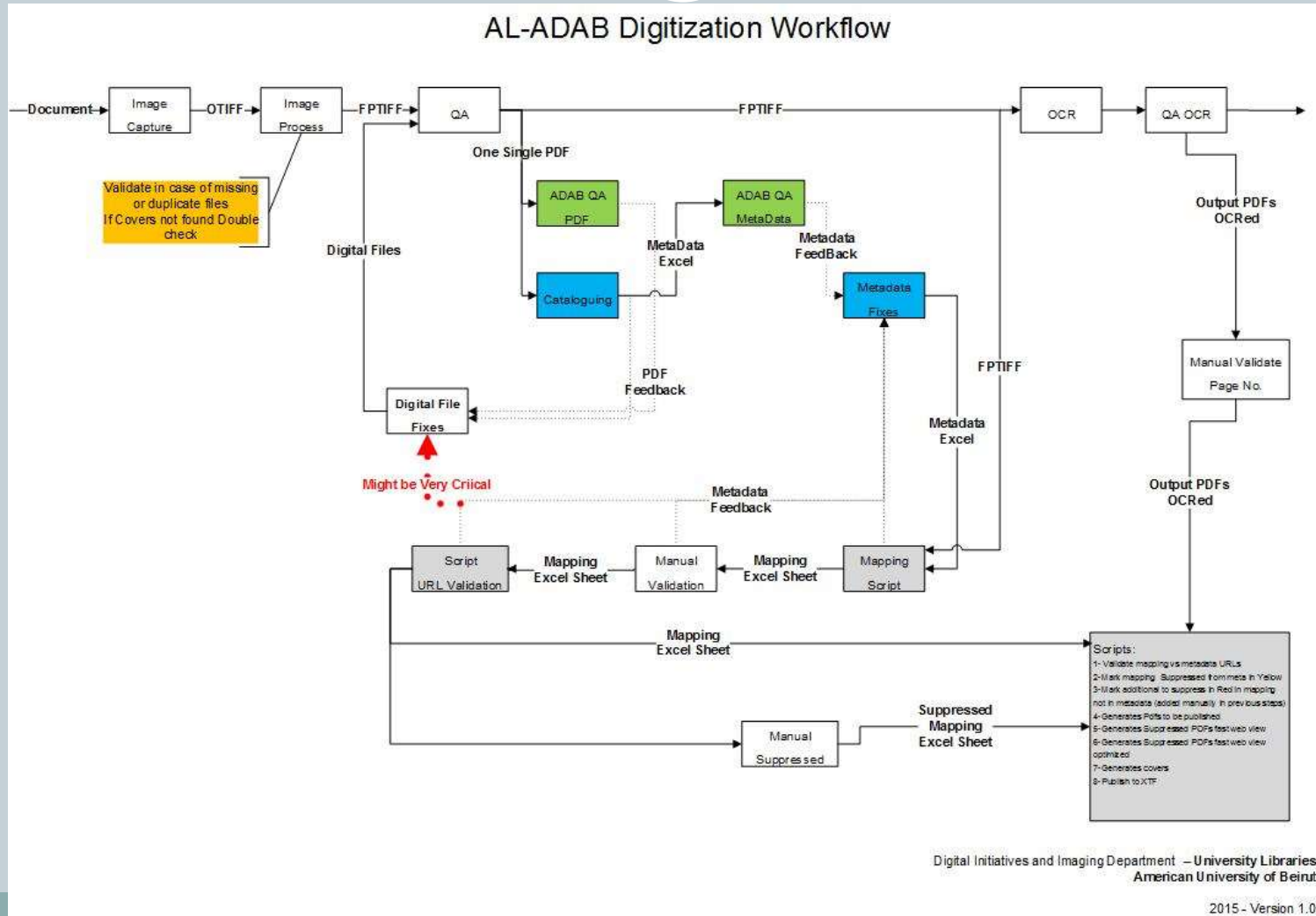
- Digitization for **access**
 - Browse / Discover
 - Index & Search
 - View / Read
 - All the above facilitates the Metadata Creation

PRE Digitization – Material Preparation



- Availability of all Volumes:
 - Secured at least **two copies** of each volume (60 volumes covering 60 years)
 - **Triage and selection** of volumes for Image capture
 - (**During Image Capture Phase**): We discovered that couple of volumes had **different content** for the same issue. The reason was mainly related to **ensorship** which lead to different publications of same issue in different geographical area.
- All Volumes are complete
 - But still missing some Covers

Digitization - Workflow



Digitization - Image Capture



- **Digitization** is the **conversion** of an **analog** source material such as document, image, sound/audio to be represented in a **numerical format** (mainly binary) which will allow access through computers and to store on different hardware medium such as Hard Disk, CDs, etc...
- FlatBed and Book **Scanners** were used to digitize the 60 volumes. It generated around **50,000** digital images.

Optical Character Recognition



OCR is a process / a computer software that converts digital images of text from pixels into encoded machine text (editable and searchable texts). **It turns a picture of text into text itself.** In other words it produces something like a TXT or DOC file from a scanned JPG or tiff format representing a printed page.

Optical Character Recognition



- **OCR Advantages:**
 - **Index** and **Search** through the generated text.
 - **Edit** the generated text
- For this initiative we are interested in the OCR to achieve **full text search**.

Optical Character Recognition



- **OCR Accuracy**

- The OCR Accuracy was calculated on **character level**. It is the percentage of characters that were correctly recognized by the OCR engine compared to the total characters in the page.
- It varied from 84% to 99% in “Al Adab” depending on the fonts. Old volumes had less accuracy than New volumes.

OCR Accuracy at the font level



OCR Results for each decade

- 1950s: 84.92% - 98.71%
- 1960s: 92.69% - 98.21%
- 1970s: 90.4% - 98.37%
- 1980s: 93% - 99.14%
- 1990s: 95.88% - 98.2%
- 2000s: 95.92%-99.4%

Large fonts better results then small fonts

Table 1. Arabic characters.

Character Name	Isolated	Connected		
		Start	Mid	End
Alif ا	ا	ا	ا	ا
Baa ب	ب	ب	ب	ب
Taa ت	ت	ت	ت	ت
Thaa ث	ث	ث	ث	ث
Jeem ج	ج	ج	ج	ج
Haa ح	ح	ح	ح	ح
Khaa خ	خ	خ	خ	خ
Daal د	د	د	د	د
Thaal ذ	ذ	ذ	ذ	ذ
Raa ر	ر	ر	ر	ر
Zaay ز	ز	ز	ز	ز
Seen س	س	س	س	س
Sheen ش	ش	ش	ش	ش
Saad ص	ص	ص	ص	ص
Saad ض	ض	ض	ض	ض
Dhaad ظ	ظ	ظ	ظ	ظ
Ttaa ط	ط	ط	ط	ط
Dthaa ظ	ظ	ظ	ظ	ظ
Ain ع	ع	ع	ع	ع
Ghen غ	غ	غ	غ	غ
Faa ف	ف	ف	ف	ف
Qaf ق	ق	ق	ق	ق
Kaf ك	ك	ك	ك	ك
Lam ل	ل	ل	ل	ل
Mem م	م	م	م	م
Noon ن	ن	ن	ن	ن
Haa ه	ه	ه	ه	ه
Wow و	و	و	و	و
Yaa ي	ي	ي	ي	ي

Optical Character Recognition



- In the old “Al-Ādāb” volumes, **different “Fonts”** were used in the same **volume (year)**, in the same **issue**, in the same **page**
- To increase accuracy, **improves images** before OCR such as de-speckle (**noise removal**) and de-skew (**straighten pages**)

التي نشرها في صحف المهجر كالسائح وغيرها. تعني عن كتب كثيرة. وقد كانت مقالاته صرخات داوية في أذن الرجعية العمياء. فقد كان

الأمين مجدداً في الوقت الذي كان فيه التجديد كفرة لا يعترف الأديب، وفي الفترة التي كانت فيها الافكار التقدمية مشاراً للسخرية والاشتراك ومطراً الشفاء...

ولم يكن أمين مشرق منجرافاً مع تيار التجديد، لأنه لم يكن من الذين ينصرفون مع تيار النزعات الفكرية من غير تفكير، ذلك انه لم يكن لعمّة شأن بعض المهرجين الذين لا يتورعون عن وضع سقط المناع في آنية التجديد الزيتف. لأن النزعة التجديدية كانت تتوحد مع دمه، بل كانت عواصف تترار في أعماقه فيطلقها براءته...

وحسبك أن تقرأ له بحمته الطويل «الداء العمياء» المنشور في كتاب «ما وراء البحار - أو النبوغ العربي في العالم الجديد» لتوفيق الرافي. فقد أوضح فيه التقيد أدق المقاييس التي يجب أن يقاس بها الأدب الحي. وقد استهل بحمته بتأييده لميخائيل نعيمة في انتقاده اللاذع لدرّة من درر شاعر الأمراء المرحوم احمد شوقي. ثم طفق يندد بالمبالغة المقتبة التي طغت على أدبنا القديم وسرت عداواها الى أدبنا الحاضر قائلاً:

«لو سئنا أحد الشعراء يرثي اسكافاً مات بين النمل والاحذية قائلاً ان الفضل مات بخرته والطمع هد ركنه والأدب امسى نيباً، ويتعجب كيف ان النجوم لم تنطق حداداً والمذهر لم يقف جثراً. أو لو قرأنا شمرأ آخر يمدح فيه انور بلشاً وحصانه الأدم بقوله: ان صيله «في قلب اوربا له ترديد». أو لو سمعنا عائقاً يتشد:

أمر بالحجر القاسي فألتمه لأن فليك فاس يشبه الحجرأ
وسألنا الثلاثة إذا كان هذا الغلو، فضحكوا منا ولا شك مشفقين لجناننا
ثم اخرج اوفهم من تحت ابطه كتاب علم الماني والبيان. وأظفر الثاني ديوان المثني أو ابن القارض. وفتح الثالث كتاب نهج البلاغة، وقدموها لنا وقد لاحت ابهامة الانتصار على تمورم ولسان حالهم يقول: تهلوا هنا قواعد البلاغة وخذود البيان ثم لا تتناجون الى سؤال...

. ثم يُنور أمين مشرق على هؤلاء ثورة مشفق ساخر قائلاً:
«هؤلاء القوم وبلا لاسف ممدورون بعض العذر. كيف لا وكل ما تلهوه منسدة ان اصبحوا يتناجون بالكلمات يتدبىء به «حدثنا سبيل بن عباد» وينتهي بنرح المعانيات السبع؟ أيام التفتيح على حفظه مثاله؟ أو ليس طبعياً ان تنمو التينة معوجة اذا ربطناها بخائط معوج؟ وهل الذئب ذئب الارض

أمين مشرق: الأديب كمنعرج

يقدم هادي طرزي

الشاعر العبقرى المجهول الذي صان مائة جبينه، وسقى دوحه شاعر ينسه بدمع عينه وعبرات فؤاده، هو أسى وأرفع قدراً

من الف أمير للشعر يغس براءه في قلوب غيره من الشعراء ثم يسامر ذوي الجاه والسلطان ليغمره بالعطايا والصدقات.

والأديب العبقرى المجهول، الذي يعترف أدبه من ينبوع الحياة، من شعبه الخائز ومن شعوب الأرض الخائرة المتألمة المستطلعة الى مستقبل أفضل، هو أسى وأرفع قدراً من أديب ينثر ما ينظمه الاخرون، ويخرفه، ثم يتماثلق هذا ويحتو امام ذلك، طبعاً في الحطام، حطام الدنيا الزائل الذي كان في الماضي البعيد والفريق لطخة عار في جبين الأدب العربي. إنه يتسلق سلم الشعوات قاصداً الشهرة، والشهرة قد تكون في بعض الاحيان كلومس من استرضائها كان دونها قدراً كما يقول ميخائيل نعيمة...

فالشهرة ليست ضرورة للأديب، بل قد تكون وبالاً عليه عندما تحيط بهالة من التقديس، ودونك ما فعلته الشهرة العريضة بكبار ادبائنا المعاصرين في إنتاجهم الرخيص...

والمرحوم أمين مشرق كان شاعراً مجهولاً وكاتباً غير معروف الا لدى هواة الأدب المهجري، ولم ينصفه أحد من أديباء العربية سوى الدكتور محمد مندور في كتابه النفيس «في الميزان الجديد».

وقد طلبت من صديقه ميخائيل نعيمة عام ١٩٥٠ أن يزودني ببعض المعلومات عن حياة هذا الشاعر الفريد، وإذا بمعلومات نعيمة ضئيلة أيضاً في هذا الموضوع فلقد كتب لي يقول:
«المرحوم أمين مشرق كان صديقي ولكني لا اعرف الكثير عن حياته، الا انه هجر الى الولايات المتحدة اولاً ومنها الى الأكوادور. ثم عاد الى لبنان عام ١٩٣٣ حيث كانت الفرصة الأولى لتعارفنا. وتزوج في العام ذاته وعاد الى الأكوادور. وبعد عام أو عامين قضى في حادث سيارة. أما آثاره الشعرية - على قلتها... فأكثرها من طراز ممتاز...»

★

لم يختلف المرحوم أمين مشرق كتاباً تقريباً ولكن مقالانه

سارت جنازة كل فضل في الوري
وتبتم الايتام أول مرة
والله ما مات الوزير وكنتمو
لما ركبت الآلة الحدباء !!
ورمى الزمان بصرفه الفقراء !
فوق التراب أعزة أحياء !!!



أما نظرة امين مشرق الى الحياة فقد كان يشوبها
شيء من ظلام التشاؤم ، إلا ان تشاؤمه هذا كان قاصراً على
نفثاته الوجدانية الذاتية لأنه كان متألماً في ديار العربية . غير انه
كان مؤمناً بادب الالتزام ، أدب التوجيه ، أدب الواقع .
وحسبك ان تقرأ له مقاله الطويل « أردية الآباء » المنشور في
كتاب « بلاغة العرب في القرن العشرين » فقد خاطب الجيل

في ضعف المجتمع والمخالة .

وهكذا تكون الثقافة المشوذة متفتحة غير ما انتجته
الإنسان في خلال جهاده التاريخي . وبدا تم ما صفة أخرى
من صفاتها الأساسية . فإني كان لكل ثقافة ميزاتها الخاصة
ووجدتها ، فان الثقافات كلها لتنتفي في ثقافة إنسانية شاملة .
ذلك ان وراء المواطن وابن المجتمع الإنسان بذاته الأهميّة :
الإنسان بنفسه الحقيقة والخرافة عنها ، بتساميه الى الأعلى
والمخادبة الى الأدنى ، بإيمانه وخوفه واطمئنانه وقلقه ، بألمه
وأمله وحزنه وفرحه ، بتجبره وانكساره ، بمجآله وقبحه ،
بإمكاناته وحدوده . هذا الإنسان قد جاهد الطبيعة ، وجاهد
نفسه بظروف وأشكال وأزمات مختلفة ، فتولدت عن هذا
الجهاد قيم جليلة تمثل في الواقع من الأدب والتصوير والنحت
والموسيقى وسواها من فنون التعبير ، وقيم عقلية تيسدو في
ما اكتشف من حقائق الوجود وفي الانتظام العقلي المؤدي الى
هذا الاكتشاف ، وقيم خلقية تنجس في ما تسمى اليه من مراقي
الحير ، وما حقق منه فرداً ومجموعاً ، وقيم روحية في تساؤله
عن نفسه وعن مبدعه وفي تعطشه الى الله فخآه اولاً ثم يلجأ الى
رحمته وغفرانه ، ويمجد خلاصه في تسليم نفسه اليه والركون الى
محبه الشاملة التي لا تاندنها محبة .

ان الثقافة المشوذة تبرز على هذه القيم ، وتسعى الى
تمثلها وتمييتها دون تردد او خشية لأنها تعي انها تولد في
مجموعها خلاصة الحضارة الانسانية ، وليس للثقافة الحية الفاعلة
ان تنكفي على نفسها ، وتنكش في صدقها ، وتقطع صلتها
بالباقي الفاعل من هذه الحضارة . انها إن فعلت ذلك اختفت
وخفت معها مجتمعا ، فلم يبق هذا المجتمع ولم يكن
مستحقاً لبقاءه .

على ان هذا التفتح الحضارة الانسانية ليس سوى خطوة
لعمل الثقافة الحية الحقيقي ، وهو مساهمتها الفعلية في هذه
الحضارة . ان هذه المساهمة هي الصفة المميزة الأخيرة للثقافة
المرجوة تتم بنتيجة الصفات الخمس السابقة التي ذكرناها ، وبها
تبلغ الثقافة غايتها وتؤدي رسالتها .

والانتاج الحضاري هو ابدأ من عمل الأفراد : اولئك
الذين يؤهلهم الإبداع استمدادهم الفطري وجهادهم العقلي
والروحي . فعلى المجتمع ان يكتشف هؤلاء الروحيين ويرعاهم ،
ويفسح لهم مجال الانتاج . ان وراء كل حضارة قلة مبدعة

من الناس : قلة تميز عن الكثرة لا بالمال ، او الجاه ،
القوة المادية ، او الزعامة الشعبية ، بل بالاستحقاق الفدا
طبيعة وكسباً ، قلة تحقق القيم وتعممها في المجتمع ، قلة ت
لا لذاتها بل للغير ، قلة لا تتعالى ولا تتجبر ، بل تحب وتح
وتعطي ، قلة متألفة ، متعارفة ، منها تنطلق قوى الة
ومجاري الانبعاث ، ومصادر الخلق والابداع .

لقد اصبح من نافلة القول ان نردد ان أثر أي مجتم
إنساني هو في قيمة إنتاجه الحضاري ، فالتاريخ أبلغ شاهد
ذلك . ولكن هذه الحقيقة لم تتأصل بعد في نفوسنا ، ولم ت
في كياننا . على اننا نفي ان تنمى الطريق الى المجتمع :
أفضل . فلتشأن فان فضل هذا المجتمع يكون بنسبة ما يحقق
هذا الانتاج يؤديه الموهوبون الجادون من ابناءنا ، فيخلد
ويخلدون ، ويبقون ويبقى معهم مجتمعنا ، على تعدد ال
والأجيال ، غذاء دائماً وكثراً مترابداً الانسانية جماء .

والآن بعد ان لمسنا الخطوط الكبرى للمجتمع ال
الأفضل ، وأنماها الصفات التي تميزها ثقافتها ، بقي علينا
تبيين الخطى المتفاضة والوجبات المفروضة لتكوين هذه ال
المشوذة للمجتمع العربي المنشود .

من هذه الوجبات ما هو ملقى على عاتق الدولة وأ
الحكم ، ومنها ما يتعلق بالشعب عموماً ، ومنها ما يسأل
رجال الثقافة بصفة خاصة .

اما الدولة ومن دار في قلبها الاعلى من الغائبين على ا
او الطامعين فيسه فواجبهم الاول صيانة حرمة الثقافة
اذ الثقافة اصيل وارقيم وانفس من ان تكون وسيلة لاية . العقل ، وا
والشخصية الانسانية المنطوية على امكانات الحرية والمسؤولية والكرامة -
هي غاية الغايات : لا تشبهان بل تحترم ، لا تستخدم بل تحترم ، لا ت
يركع على عتبتها . ان الحاكم او السياسي الذي يتخذ من التربية والذ
اي شأن آخر من شؤون الثقافة سبباً لتحقيق غرض شخصي او حد
كان يسمى لتأجيل مهمة التعليم ان هو غير جدير بها ، او لاستغلال الما
لتسكين نفوذه وعجارية اعدائه ، او لاقالة الطلاب في سبيل غاية حز
مطمع على ، او كان يبعد لاتباع الاقلام من اجلس شهرة ، او
المراهب وتعثيرها - ان من هذا الحاكم او السياسي يظن امته في ال
ويغرض مرخ بناها باعائه جوهر الفكر والثقافة ، وتدنيه جرماً
والروح . فثالث هذه القيم عن ان تكون عطية لطامع ، او الدولة يبد
ساحر ، وساء قال امه ينط ارباب الشأن بها الى هذا المردك !
ولنا في تاريخنا العربي اسوة حسنة . هو مليء بالخيار النجدة والرعاية
كان يبذلها الخفاء لارباب العلم ومؤسسه .

وينتج عن هذا الاعتبار واجب الدولة الثاني : وهو

. ناصــــر الــــريــــاط .

اللغة العربية هي الوشيجة الأولى التي
تصل بين كل العرب، وريما الوحيدة
التي تحتوي كل ما تمثله العروبة

غيرها على الرغم من الدلائل التاريخية
والجغرافية المغايرة، فمآذا نستنتج من
ذلك؟ هل العروبة كما نعرفها اليوم،
وبحكم موطن نشأتها ونموها، سليل
الكلاسيكية؟ أم ان بعض العرب الذين
لم يشاركوها في الحضارة الكلاسيكية
أكثر عروبة من غيرهم الذين انتموا
إليها في وقت ما؟

بالطبع لا. بالإضافة إلى ذلك، ما الذي
يؤكدنا ان نستشف من دراسة تراث
البلاد العربية الأخرى التي لم تتشارك
مع حوض المتوسط في تراثه
الكلاسيكي؟ هل نصنف العراق وفق
انتمائه إلى التراث الرافدي - الفارسي
مثلاً، وهل نربط اليمن بالحيشة، أو
عُمان بالهند، أو السودان بقلب القارة
الافريقية، كما تتطلب معطيات
الجغرافيا والتاريخ؟ لا توجد إجابات
وافية عن هذه التساؤلات، ولكن من
الواضح ان قضية التراث المشترك تثير
من المشاكل أكثر مما تحل.

لغة مشتركة؟

لنترك إذن هذا السؤال العويص جانباً
وننتقل إلى اللغة العربية، التي يمكن
اعتبارها بحق الخاصية الأساسية
المشتركة بين العرب، ووعاء هويتهم
الحضارية وتاريخهم المشترك بشقيته
المثبت والمأمول. فعلى الرغم من تباين
وتعدد لهجات أبناء الدول العربية، فإنهم
مازالوا يكتبون ويقرأون ويفكرون بلغة

العربية المعاصرة يتقاطع ويتداخل مع
تراث دول أخرى إسلامية في غالبيتها،
مثل إيران وتركيا وإسبانيا ووسط آسيا
وبدل الصحراء الافريقية خلال الفترة
الإسلامية كلها وحتى بداية القرن
العشرين. بل إن هذا التدخل قد
استمر حتى بعد احتلال الاستعمار
الأوروبي لغالبية أراضي السلطنات
الإسلامية في العصر الحديث، وبعد
نشوء الدول الحديثة بحدودها الجديدة
وأعلامها الجديدة وأناسيدها الوطنية
الجديدة وجارات سفرها الجديدة.

فإذا ما نظرنا إلى ما قبل الإسلام فما
الذي سنجد؟ نجد ان بعض البلاد
العربية تتشارك مع معظم دول حوض
البحر الأبيض المتوسط في التراث
الكلاسيكي، فقد عاشت بلاد الشام
ومصر وشمال أفريقيا ألف سنة (بين
حوالي ٣٣٠ قبل الميلاد و٦٤٠ بعد
الميلاد عموماً) كجزء من عالم متوسطي
كلاسيكي، هيلينستي بدءاً، وروماني
لاحقاً، ومسيحي بيزنطي في النهاية.
وقد تفاعلت المنطقة مع تراثها ذلك:
فاخذت منه وأغنته، وتمامت معه،
وساهمت في العديد من منجزاته
الثقافية والأدبية والعلمية والفنية
والمعمارية. بل ربما كان أبناء المنطقة
الساغ الاطول في دفع عجلة الحضارة
التي اصططلحت أوروبا على تسميتها بـ
«الحضارة الكلاسيكية»، والتي اعتبرتها
مهذب حضارتها هي نفسها من دون

To improve search results and indexing!

To compensate the low accuracy of OCR for old volumes!

To simply complement the OCR phase and to complete the digitization process!

To guarantee the long term preservation and accurate retrieval of digital objects (articles of al-Adab)

METADATA

Al-Ādāb Magazine Archives



CATALOGING AND INDEXING

OR

DESCRIPTIVE METADATA

Cataloging and Indexing



- **Cataloging of al-Adāb magazine** is the **description** of the content at the **article level** (Author, Title, pages, and source, etc.) using a bibliographic format (Citations)
- **Indexing** is the Subject Cataloging or the description of **subjects** of the **articles** (subject headings or controlled vocabularies or keywords)
- This cataloging process is called **Descriptive metadata** for the digital objects

Cataloging tools



- Cataloging of the magazine is processed at the article level keeping the hierarchy
 - Magazine top level
 - Issue level
 - Article level
- Use of the Library system for analytical cataloging (Multilevel)
- Use of Anglo-American Cataloging Rules for Descriptive cataloging (AACR2)
- Use of Library of Congress Authorities
- Use the Arabic language in Cataloging and Indexing
- Use of MARC 21 Format

Indexing process and tools



- Use of Library of Congress Subject Headings (translated into Arabic by University Libraries)
- Detailed indexing in Arabic (2-70 Subject headings or keywords for conferences and meetings papers)
- Search could be at the Library Catalog OPAC, but the use of the Content Management System (XTF) of the AUB Libraries enables the full text search.

Metadata Export



- Export the Descriptive metadata (Cataloging and indexing field) to multi format:
- **DC 15 elements (Dublin Core) metadata**
- **MARC Exchange format** (ISO 2709)
- **MARC 21** standard
- **CSV** (Comma Separated Value)

Exported Metadata Fields 1



- System Record Number (Unique Identifier)
- Title of the article
- Main Author
- Added Authors
- Pages
- Source that includes:
 - Title of the magazine at that time
 - volume no.
 - issue no.
 - Date

Metadata Fields 2



- URL
- Rubric or Category of the Subject as categorized by the Magazine أبواب المجلة
- Subjects (LCSH) 2-70 subjects for each article (pre-coordinated vocabularies)
- The Personalities (Personal Names as subject) 600 field in MARC 21

Subject headings examples



- **Subjects Headings or Keywords in Arabic (الكلمات المفتاحية)**
 - الرسامون—اسبانيا—مقابلات
 - سوريا—الحياة الفكرية—القرن العشرون
 - الادب العربي في المهجر—تاريخ ونقد
 - اللغة العربية—اللهجات—لبنان
 - البعلبكي، منير، 1918-1999. فلسفة من الصين
 - القصة الاميركية -- ترجمات الى العربية
- **Personal authors as subject (الشخصيات كموضوع)**
 - السكاكيني، خليل، 1878-1953
 - قرم، جورج، 1896-1971
 - قرم، جورج، 1940-

Metadata mapping to CMS



- **Metadata** could be **mapped** to any Content Management System (**CMS**) from the MARC file such as: Drupal, WordPress, Dspace
- **Keywords** could be **tagged** in a **Content Management System** thesaurus or taxonomy or controlled vocabularies
- Full text can also be mapped and included in the **Content Management System** . Full text could be indexed and retrieved due to the OCR technique that transforms Arabic image text into searchable text.
- Search engine in CMS can retrieve full text and metadata in “Simple Search” and “Advanced Search”

Full Text SEARCH



- Upload all the information such as images, text, metadata into a full text searchable system.
- Provide a web interface that allows access to the digitized magazine.
- Provide a viewer/reader feature

XTF



- **XTF:** (<http://xtf.cdlib.org/>)
 - Powerful open source platform for providing access to digital content.
 - Developed and maintained by the California Digital Library.
 - Ability to create indexes on any XML element or attribute; entire presentation layer is customizable via XSLT.
- XTF provides out-of-the-box support for the following types of documents:
 - Microsoft Word, PDF, Web pages (html/htm), XML, encoded plain text, Scanned books from Internet Archives and HathiTrust
- Based on Lucene: An open free software that provides full text indexing
- Major “out-of-the-box” feature: User interface with search/browse and document views

Al-Ādāb Magazine Archives



CHALLENGES

Challenge of Censorship



- 2 editions for the same issue due to the Censorship in some Arab Countries:
 - One for standard distribution
 - One for distribution in designated country
- Looking for the complete edition (Missing issues and missing pages)
- Number of pages varies according to the omission of articles in the censored edition (affecting numbering in digital files and in the metadata)

Challenges for metadata



- Censorship: Problem at the citation level for the issue of the second edition (2 different citations for the same article)
- The periodicity / Frequency changed from monthly to bimonthly to quarterly during the life of the magazine
- Changing in the rubrics of the magazine (أبواب المجلة)
- Additional Citations for the continuations (التتمات) and combining the pdf files in one link
- Add more subject headings to compensate bad recognition in ocring of Arabic printed letters

Challenges at Technical level



- Paper quality (yellowish and fade in some issues in the 1950s)
- Binding margins require additional touching
- Covers are missing due to binding mistakes
- Printing quality varies

Challenges in Image Capture



- Image capture in Black & White for the text
- Image capture in grey scale for the images inside the text (Re-do)
- Image capture in color for the covers

Challenges for Arabic OCRing



- Arabic language is cursive language where letters are attached and each letter change according to its position in the word
- Arabic letters are overlapped in the 1950s
- Different font types and sizes in the same page
- Titles printed in calligraphic fonts
- Accuracy rate of recognition was calculated at the font level.
- Use of Sakhr Automatic Reader for Arabic OCRing that enables teaching new letters

Challenges for Arabic OCRing (cont.)



- ◉ Some characters have the same form and are only distinguished by the position of various dots relative to the main character block. Given that dots are considered as noise and OCR tends to remove them.
- ◉ Space between two connecting Arabic characters can vary in size and shape and use of kashida (extended letter) for alignment
- ◉ Vowels: (حروف العلة الطويلة (ا ، و ، ي) are different from diacritical marks or short vowels (حروف العلة القصيرة) examples fatha, dhammah, kasrah (not consistent)

وهدف المجلة الرئيسي ان تكون ميدانا لفئة اهل العلم الواعين الذين يعيشون تجربة عصرهم ، ويُعدّون شاهداً على هذا العصر : ففيهم يعكسون حاجات المجتمع العربي ، ويعبّرون عن شواغله ، يشقون الطريق أمام المصلحين ، لمعالجة الاوضاع بجميع الوسائل الحديثة . وعلى هذا ، فإنّ الادب الذي تدعو اليه المجلة وتشجعه ، هو ادب « الالتزام » الذي ينبع من المجتمع العربي ويصب فيه .

والمجلة ، اذ تدعو الى هذا لأدب الفعال ، تحمل رسالة قومية ملي . فتلک الفئة الواعية من الادباء الذين يستوحون أدبهم من مجتمعاتهم يستطيعون على الأيام ان يتحققوا جيلاً واعياً من القراء يتحسسون بدورهم واقع مجتمعاتهم ، ويكونون نواة الوطنيين الصالحين . وهكذا تشارك المجلة ، بواسطة كتابها وقراءتها ، في العمل القومي العظيم ، الذي هو الواجب الأكبر على كلّ وطني .

على ان مفهوم هذا الأدب القومي سيكون من السعة والشمول حتى ليتصل اتصالاً مباشراً بالأدب الانساني العام ، ما دام

وؤاد الشايب	علي ادهم
عبد الله عبد الدائم	ذو النون ايوب
مارون عبود	منير البعلبكي
عبدالله العلابي	خليل تقي الدين
توفيق يوسف عواد	شكيب الجبري
نبيه امين فارس	جورج حنا
شكري فيصل	شاكر خصباك
نزار قباني	رئيف خوري
صباح محي الدين	عبدالعزیز الدوري
انور المعداوي	قسطنطين زريق

الذي ينبع من المجتمع العربيّ ويص

وللمجتمعات ، اذ تدعو الى هذا لأ

مالي . فتلك الفئة الواعية من الادباء

يستطيعون على الأيام ان يخلقوا



Cover of the **last issue** from the **last publication year 2012**



University Libraries



Thank you

<http://www.aub.edu.lb/ulibraries/>

Basma Chebani (email: bc01@aub.edu.lb)
Head of Cataloguing And Metadata Services, University Libraries

Elie Kahale (email: ekahale@aub.edu.lb)
Head of Digital Initiatives and Imaging Department, University Libraries