CHAKEL-DB: Online Database for Handwriting Diacritic Arabic Character

Houda Nakkach[a], Sofiene Haboubi[b] and Hamid Amiri
LR-11-ES17 Signal, Images and Information Technology (LR-SITI-ENIT), Tunis El Manar University, National Engineering School of Tunis 1002, Tunis Le Belvédère, Tunisia

Keywords: CHAKEL-DB, Handwritten Arabic Character, Diacritical Marks, Online Character Recognition.

Abstract: Our paper presents an online database for handwriting Arabic characters, so-called "CHAKEL-DB": with "chakel" means diacritic in Arabic language. The database contains 3150 collected samples that present alpha-numeral characters with diacritical marks. The data are collected from more than 68 writers having different origins, genders and ages. The data are available in character and stroke levels. We built an elementary recognition system to test our database and to manipulate a large vocabulary and a huge quantity of variation style in the collected data. "CHAKEL-DB" is available for the purpose to improve handwriting research field, to facilitate experiments and researches. This database offers files in XML format.

1 INTRODUCTION

With the increasing spread of hand-held devices such as PDA, tablet-PC and Smart-phone, the handwriting is introduced as a new human-computer interaction modality. As result, an increased demand for a high performance of the on-line handwritten recognition system, which becomes a popular and fascinate field for researcher in the recent years. On-line handwritten recognition is presented by a sequence of strokes in 2D (x, y) coordinates form, which are the result of succession points obtained from the pen-down/pen-up signals. Such data is known as digital ink and can be considered as a dynamic representation of handwriting. An on-line handwritten recognition system typically includes three main elements: An output display adjacent to a touch sensitive surface, a database of characters or general gestures and a recognizer system as a software application which interprets the movements of the stylus across the writing surface and translating the resulting strokes into digital character based on database samples.

In this context, more literature works have been developed for the Latin characters case, but few of them for the Arabic case, challenged by the various nature and unconstrained cursive Arabic Script (Nakkach et al., 2016) (Biadsy et al., 2011). Diacritical marks are marking which are written either above or below a letter. The diacritics "Fatha", "Damma", and "Kasra" indicate short vowels. "Sukun" mark indicates a syllable stop, and "Fathatan" indicates nunation ("tanwin" or double vowels) that can accompany "Fatha", "Damma", or "Kasra". The three vowel letters, which are "Alef", "Waw" and "Ya" are used to indicate long vowels. The "chadda" vowel represents doubling (or gemination) of a consonant (Table 1). So, along with the dots and other marks representing vowels, this makes the effective size of the Arabic alphabet about 2040 characters (equation 1). In fact, we have 23 characters with 4 forms, 5 characters with only 2 forms and we have 20-mark models for each form of character.

$$(23*4+5*2)\ *20=2040 \quad (1)$$

Table 1: Exhaustive list of "BA" character diacritical models.

Nunation & chadda	Long chadda	Short chadda	Nunation	long vowel	short vowel	Sukun	Main body	With
-	-	بَ (bb)	-	-	-	بْ (b)	ب (BA)	None
بًا (bban)	بَا (bbaa)	بَ (bba)	بًا (ban)	بَا (baa)	بَ (ba)	-	-	fatha
بٌ (bbon)	بُو (bboo)	بُ (bbo)	بٌ (bon)	بُو (boo)	بُ (bo)	-	-	damma
بٍ (bbin)	بِي (bbii)	بِ (bbi)	بٍ(bin)	بِي (bii)	بِ (bi)	-	-	kasra

[a] https://orcid.org/0000-0003-1699-3812
[b] https://orcid.org/0000-0001-5270-3830

Nakkach, H., Haboubi, S. and Amiri, H.
CHAKEL-DB: Online Database for Handwriting Diacritic Arabic Character.
DOI: 10.5220/0008804307430750
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 743-750
ISBN: 978-989-758-402-2; ISSN: 2184-4321

In addition to these usual issues for the handwriting recognition, recognizing Arabic character needs to deal with ligatures, diacritics, multi-variability writer style, bad writing habits like touching characters, misplacement of dots, etc. Therefore, most research in the character recognition field is moving towards experimentation of their proposed system to test new approaches and algorithms. In this case, they need database to validate their theories, for that they use databases available for on-line and off-line writing or they develop their own database as presented in (Plamondon et al., 2000) (Tagougui et al., 2013) (Vinciarelli, 2002)(Jäger et al., 2001)(Steinherz et al., 1999).

On other wise, examining the existing databases, we find that none of them deals with on-line / off-line cursive Arabic writing with diacritical marks. Thus, the development of a new database is useful for the scientific community, with the continuous grow up of interested searchers number work on the recognition of Arabic writing field. To respond to this need, we propose in this paper a new database for the Arabic diacritical characters named "CHAKEL-DB".

So, the remainder of this article is organized as follows: in section 2., we begin with an overview of the related works. Section 3. presents the data preparation and acquisition stage. Supplemented by, a quick overview of the existing data formats and databases, presented as a comparative study based on indicators reflecting the requirements to choose the formats to be adopted in our work. The experimental results of our work are presented in section 4. And we close the paper with a conclusion in section 5.

2 RELATED WORK

All scientific domains have several standard databases for developing, evaluating and comparing different techniques developed for their various tasks.

The field of recognition of handwriting is not an exception; in this context, the handwriting recognition community has proposed many databases, some of them will be listed below, to present some statistics and comparisons based on a few specific dimensions of Handwriting. These dimensions allow us to make a subdivision in the field of handwriting. We can divide handwriting recognition into two fields, depending on the form in which the data is represented on-line or off-line.

In the case of on-line recognition of handwriting, the user must write on a digitizing tablet with a special stylet, so that the lines are sampled by the coordinates (x, y) of the spaced time intervals. However, in the case of off-line handwriting recognition, the user writes on a paper which is then scanned by a scanner. In this case the data is presented to the system as an image, which requires segmentation to binarize it through the threshold technique based on the color pattern (color or gray scale), so, the image pixels are either 1 or 0.

The on-line case concerns a spatial-temporal representation of the input; while the off-line case, involves an analysis of the spatial luminance of image.

The most important and widely used handwritten databases include:

- IAM databases: used as collections of handwritten samples; they are adopted for a variety of segmentation and recognition tasks. Several off-line and on-line databases have been developed within the IAM, such as:
- IAM-DB (IAM handwriting database) (Marti and Bunke, 1999): this handwritten database proposed since 1999, contains forms of unconstrained western handwritten English text. The IAM Handwriting Database 3.0 published in 2002 includes contributions from 657 authors making a total of 1539 handwritten pages including 5685 sentences, 13353 lines of text and 115 320 words. The database is labelled at the level of sentences, lines and words, and it has been widely used in word tracking, writer identification, text segmentation Handwriting and off-line write recognition. This database is presented by image files described by meta-data files in XML format.
- IAM-OnDB (Liwicki and Bunke, 2005): The database is a collection of handwritten samples on a white-board acquired with the E-Beam system. The data is stored in XML format which, in addition to the transcription of the text, also contains information and demographic data about writers. The database includes 221 authors contributing a total of more than 1,700 forms with 13,049 lines of labelled texts and 86,272 occurrences of words from a dictionary of 11,059 words. In addition to the recognition of on-line writing, the database was also used for on-line writer identification and gender classification from handwriting. The collected data is stored in XML format, and available in tif-images.
- Rimes (Augustin et al., 2006): it is an off-line database composed of emails sent by individuals to companies or administrations. It contains 12,723 pages corresponding to 5605 emails of 1,300 volunteers. Collected pages scanned and annotated in RIM format, database is fully used

for the evaluation of tasks such as document layout analysis, handwriting recognition, identification and verification of writers, identification of the logo and the extraction of information.

- NIST: The National Institute for Standardization and Technology (NIST) developed a series of databases (Wilkinson et al., 1992) containing handwritten characters and numbers supporting tasks such as field isolation, detection and removal of boxes in forms, segmentation and character recognition. The NIST Special Database 1, includes samples provide by 2100 authors. The latest version of the database, the Special Database19[1], includes handwritten forms of 3600 writers with 810,000 images of individual characters with their meta-data. This database has been widely used in a variety of handwritten figures and character recognition systems.
- MNIST[2]: An off-line database for handwritten digits. It is a subset of a larger set available from NIST, includes a training set of 60,000 examples and a test set of 10,000 examples. This database has been widely used in machine learning and in several digit recognition systems.
- CEDAR CDROM1 (Cheriet et al, 1994): An off-line database contains 5632 city/state handwritten words, 4938 handwritten states, 9454 postal codes. Which presents a total of 21,179 manuscript digits and 27,837 alphanumeric characters. This database has been used for the evaluation of several systems including handwriting segmentation, recognition of cursive digits character recognition and word segmentation. A CD-ROM 2 contains machine-printed Japanese character images.
- IRESOFF IRONOFF (Viard-Gaudin et al., 1999): A handwriting on/off database, for each character or handwritten word, both online and offline signals are available. It consists of 4086 single digits, 10,685 isolated lower-case letters, 10,679 isolated capital letters and 31,346 words from a 197-words lexicon (French: 28,657 and English: 2 689).
- UNIPEN[3] ((Guyon et al., 1994): Proposes an online database for Western handwriting, since 1993, contains more than 5 million Western characters, written by more than 2,200 writers.
- IFN/ENIT (Pechwitz et al., 2002): an offline Arabic handwritten words database was developed from the contributions of 411 volunteers each filling a specific form. The database presents a total of 26,400 words (city / city names) corresponding to 210,000 characters. And the meta-data includes information about the sequence of character shapes, baseline, and author.
- CENPARMI (Alamri et al., 2008): an offline database for Arabic handwriting recognition. It consists of single digits, letters, strings and words, written by 100 participants from Canada and 228 participants from Saudi Arabia. The database has been used for the recognition of Arabic characters, digits and word marks.
- ADAB (Abdelaziz et al., 2014) (Abed et al., 2009): an on-line database proposed in 2009. It contains Arabic words corresponding to the names of Tunisian towns and villages, manuscripts by more than 173 different authors, totalling more than 29,922 Arabic words and 157,792 characters.
- AltecOnDB (Abed et al., 2010): An on-line handwriting database. The collected samples are complete sentences that include numbers and punctuation marks. Samples handwritten by more than 1000 writers with different origins, genders and ages. The collected data is used in various tasks such as annotation and verification, recognition, and segmentation tasks.
- OHASD (Elanwar et al., 2010): An on-line database of Arabic handwritten sentences, the dataset includes 154 paragraphs written by 48 writers aged 24-40 years of both sexes. It contains more than 3,825 words and more than 19,467 characters. The dataset is used in a word retrieval system and extraction of lines of text, as well as word segmentation and annotation.

3 OUR CONTRIBUTION

3.1 Handwriting Database Recommendations

Our contribution consists to propose a new database for diacritical Arabic character recognition systems, named "CHAKEL-DB". To achieve this goal, there are some recommendations to be respected, whether for the data structure or for the storage structure. Such as: expressiveness, accessibility, parsability, serializability, well defined and well formed.

[1] https://www.nist.gov/srd/nist-special-database-19
[2] http://yann.lecun.com/exdb/mnist/
[3] http://hwr.nici.knu.nl/unipen

Following our comparative study conducted on some standard formats (Table2), the most recommended data formats in our context are the inkML[4] and UPX[5] which are in an XML format. The inkML standard presents a set of tags that describing the electronic trajectories of writing, hence the name Ink and considering the medium level of expressiveness we plan to complete it with the UPX standard which resembles the hwDataset standard (Bhaskarabhatla and Madhvanath, 2004) with the meta information and the annotations of the presented data and in addition, it can detail several levels of hierarchy for the trajectory unlike the hwdataset which presents a single level.

3.2 Data Collection Scenario

For the collection of samples, we adopted the scenario presented by the Figure 1, which consists in the acquisition of the data, the collection of the trajectory points characteristics, the capture of the character image, and finally the annotation and verification of the data to feed our database CHAKEL-DB.

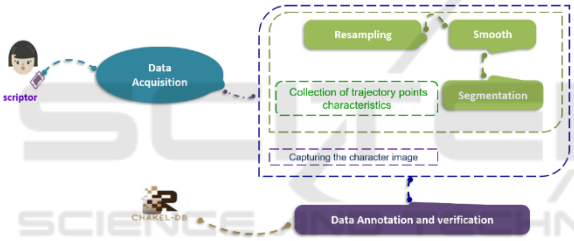

Figure 1: Data collection scenario.

For the acquisition of the character we will need of course a tactile surface. We qualify the trajectory acquired by three events (Figure 2): Action_down which indicates the beginning of the writing of the trace, Action_move which represents the movement of the stylet and the action_up which indicates the end of the trace.

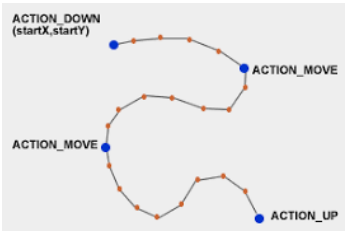

Figure 2: Event Recovery.

We characterize each point of the trajectory by its coordinates (X, Y), the force F of the writing and the timestamps. The result vector goes through a resampling step and a smoothing step. Capturing the result character image is useful for various uses at online and even offline diacritic Arabic character recognition systems. The method used for framing the character image is based on the selection of the endpoints (Figure 3).

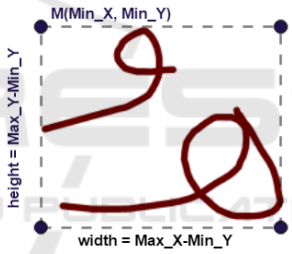

Figure 3: Framing the character image.

The concepts (Figure 4 and 5) highlighted in our sample database are the writer who defines himself by

Table 2: Handwriting Data Formats overview.

Data Format	Binary Format	Requirements					
		Expressivity	well-defined	Storage Structure			Data structure
				Accessibility	Well-formed	Parsability	Serialisability
Microsoft Ink Serialized Format (ISF)	Yes	Low	Low	Low	Low	Yes	Medium
InkML	No	Medium	High	High	High	Yes	High
Standard UNIPEN Format	No	Medium	High	Medium	Low	Yes	High
HandWriting Dataset(hwDataset)	No	High	High	High	High	Yes	High
UNIPEN XML Format (UPX)	No	High	High	High	High	Yes	High

[4] http://www.w3.org/TR/2006/WD-InkML

[5] http://unipen.nici.ru.nl/upx/index.html

his gender, his age, Hand writing, level of education, region, type of material used and level of use, type of script as well as the level of mastery of the script.

A writer can write several characters, each character identified by an unique identifier that presents his class, it is composed of several traces each one is characterized by its identifier, its duration and the speed of writing and it comes in a set of points.

Our database considers 21 diacritic models for each of the 28 Arabic characters with the 10 digits, we will have at all: 598 models.

We characterize each point of the trajectory by its coordinates (X, Y), the force F of the writing and the timestamps. The result vector goes through a resampling step and a smoothing step. Capturing the result character image is useful for various uses at online and even offline diacritic Arabic character recognition systems. The method used for framing the character image is based on the selection of the endpoints.

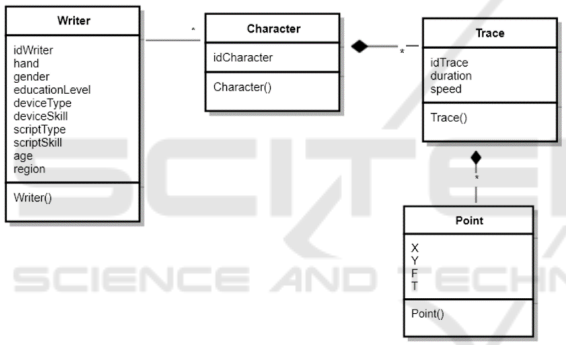

Figure 4: Chakel-DB concepts.

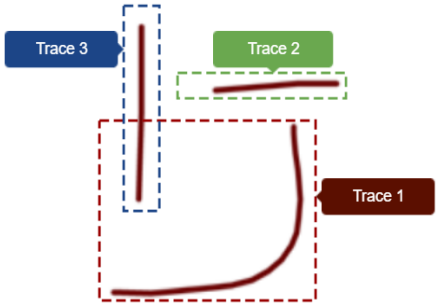

Figure 5: Character traces example.

3.3 Experimentation and Evaluation

Our experimental task for the evaluation of this database is in three stages (Figure 6): The first step is to implement the mobile frame that presents the sample collection interface, their processing to generate a set of JavaScript Object Notation (JSON) data that will be sent to the dedicated server. The data retrieved at the server level by a web services framework, which serves as a parser for JSON messages, to generate the inkML, UPX, and tif-files that feed our database. The second step is the establishment of an Arabic character recognition system to test the performance of our database. Finally, the performance test phase.

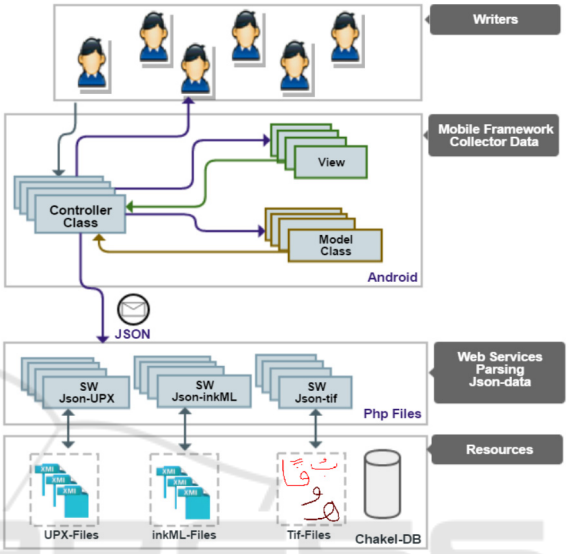

Figure 6: Architecture of the mobile sample collection framework.

3.3.1 Examples of Samples

As example of samples from our database, I present this list of UPX files(Figure 7) with their complementary inkML (Figure 8) which presents the characters figured in the tif-images of figure 9. Our data base presents also a set of binary images (Figure 10).

Figure 7: UPX Files.

Figure 8: inKML Files.

An example that presents zoom on the manuscript character online “Ha” (Figure 11) with an extract of

the character metadata as inkML file (Figure 12) and an extract of the Ha character metadata at the UPX file level (Figure) which refers to the tif-image of the character as well as the inkML complement.

Figure 9: tif images.

Figure 10: Binary images.

Figure 11: tif-image of the diacritic character "Ha".

```
<?xml version="1.0" encoding="UTF-8"?>
<ink xmlns="http://www.w3.org/2003/InkML">
  <context id="c1" idBrush="PenA1" idInkSourceRef="KTU84P" idTraceFormat="XYFT"/>
  <brush color="#FF660000" id="PenA1" width="8"/>
  <inkSource id="KTU84P" manufacturer="Acer" model="B1-750" serialNo="45041085369" description="vespa2">
    <sampleRate uniform="true" value="200"/>
    <activeArea size="1" height="889" width="744"/>
  </inkSource>
  <definitionList>
    <traceFormat id="XYFT">
      <channel name="X" type="decimal"/>
      <channel name="Y" type="decimal"/>
      <channel name="F" type="decimal"/>
      <channel name="T" type="integer"/>
    </traceFormat>
  </definitionList>
  <trace id="id001" duration="1088.0" type="penDown" speed="188.8">547.25 357.55 1.0 1542708.0,579.25 391.15 1.0 1542889.0,587.95 402.1 1.0 1542906.0,592.1 411.15 1.0 1542923.0,598.85 431.85 1.0
  <trace id="id002" duration="196.0" type="penDown" speed="10.3">955.6 945.45 1.0 1544221.0,911.25 890.9 1.0 1544241.0,866.85 836.35 1.0 1544258.0,822.45 781.8 1.0 1544275.0,778.05 727.25 1.0 154
</ink>
```

Figure 12: The metadata of the character "Ha" at the level of the inKML file.

```
<?xml version="1.0" encoding="UTF-8"?>
<UpxType xmlns="http://www.w3.org/2003/InkML" xmlns:xmi="http://unipen.nici.ru.nl/upx" xmi:version="0.9.5" version="2.0">
  <dataSetInfo name="CHAKEL-DB : On-line DataBase to Handwriting Arabic Character" contact="houda.nakkach@gmail.com" category="On-line" version="1.0" source="KTU84P">
    <dataInfo contentDesc="Data base of Handwriting Arabic Character" quality="high" style="high" numWriters="1"/>
  </dataSetInfo>
  <dataSetDefs>
    <labelSrcDefs>
      <labelSrc desc="The label is not yet defined" id="labelsrc1" name="Undefined"/>
      <labelSrc desc="A recognizer for Arabic handwriting Character" id="labelsrc2" name="Recognizer"/>
      <labelSrc desc="The label is verified by a human" id="labelsrc03" name="Human"/>
    </labelSrcDefs>
    <writersdefs>
      <writer id="id1413782619">
        <personal age="36" educationLevel="Doctoral Degree" gender="Male" hand="right" profession="teacher" region="Tunisia"/>
        <skillScript scriptType="high" value="cursive"/>
        <skillDevice deviceType="Tablet" value="high"/>
      </writer>
    </writersdefs>
  </dataSetDefs>
  <hwData id="HD1" desc="high">
    <imgInfo imgSrc="1478528743.tif"/>
    <hLevel id="level1" level="character" writerRef="id1413782619">
      <label id="Ha" labelType="CHAKEL/ENIT" timestamp="1478528743" labelSrcRef="labelsrc03">
        <alternate value="&amp;#x6be;&amp;#x64e;"/>
      </label>
      <hwTraces>
        <traceView id="1478528743.inkml"/>
      </hwTraces>
    </hLevel>
  </hwData>
</UpxType>
```

Figure 13: The metadata of the character "Ha" in the UPX file.

3.4 CHAKEL-DB Statistics and Comparison

In this section, we list some statistics about the collected data in our online database for diacritical Arabic characters. We collected handwriting samples from 60 different writers composed of men and women from various professional cadres, qualifications and ages. We were interested in keeping track of the writer's gender, age and whether they were right-handed or left-handed. CHAKEL-DB database presents 70% of data to training set, so for this set we have 55 writers, 2205 Characters, and 12238 traces.

Table 3 presents a comparison database with the other databases of the literature based on the: content level sample numbers, file type and tasks supported. We remark that: (i) our database deals with the case of the diacritic sign, the others do not. (ii) it complies with the recommendations of the online databases, and (iii) It can be used for any application field for CRS. The only limit mentioned is the number of samples that we expect to increase soon.

4 CONCLUSION

In this paper we have introduced CHAKEL-DB, an on-line database for handwriting. One of the main objectives of this project is to collect a huge database of handwritten characters from several writers, presented by mobile application as input in real-time. Unlike currently available Arabic databases, our database: (i) includes high coverage for all possible diacritical marks of the Arabic language. (ii) Data are collected from many writers with different ages and work experiences. (iii) The collected samples are available in two levels: characters and traces (iv) and writing involves a multi-colored, multi-sized brush. CHAKEL-DB contains 3150 patterns of diacritic characters and 17095 traces. For each element of the database, a corresponding meta-data file is available in an inKml format and an upx format. The database is available on-line for various research applications such as on-line recognition of Arabic handwritten diacritical characters, on-line identification/verification of writers, on-line handwriting segmentation, gender classification and on-line color identification. CHAKEL-DB will be great help and value for the research community, and can be downloaded by requesting a copy from LR-SITI Lab.

REFERENCES

Abdelaziz, I., Abdou, S., 2014. *AltecOnDB: A Large-Vocabulary Arabic Online Handwriting Recognition Database*. arXiv preprint.

Abed, H. A., Margner, V., Kherallah, M., Alimi, A. M., 2009. *Online Arabic Handwriting Recognition Competition.* ICDAR.

Abed, H. A., Kherallah, M., Margner, V., Alimi, A., 2010. *On-line Arabic handwriting recognition competition, ADAB database and participating systems*. International Journal on Document Analysis and Recognition (IJDAR), Springer, 14, 15-23.

Alamri, H., Sadri, J., Suen, C., Nobile, N., 2008. *A novel comprehensive database for arabic offline handwriting recognition*. In Proceedings of the 11th Intl. Conference on Frontiers in Handwriting Recognition.

Augustin, E., Brodin, J. M., Carre, M., Geoffrois, E., Grosicki, E., Preteux, F., 2006. *RIMES evaluation campaign for handwritten mail processing*. In Proceedings of the Workshop on Frontiers in Handwriting Recognition.

Bhaskarabhatla, A. S., Madhvanath, S., 2004. *An XML Representation for Annotated Handwriting Datasets for*

Table 3: CHAKEL-DB database characteristics.

	ADAB (Abed et al., 2009)	OHASD (Elanwar et al., 2010)	AltecOnDB (Abdelaziz et al., 2014)	CHAKEL-DB
Level of Contents	Word (names of cities), Character, Trace, No diacritics	Sentence, Word, Character No diacritics	Sentence, Word, Character No diacritics	Character, Trace, Diacritic signs
Number of samples	173 writers, 29,922 words, 157.792 characters	1000 writers, 152,680 words, 6444,530 characters	1000 writers, 152,680 words, 6444,530 characters	68 writers, 3150 Characters, 17095 traces
File Type	Upx, inkml, tif-image	Txt, isf, jpg-image	dhw, relational tables (SQL DBMS)	Upx, InkML, tif image, binary images
Supported tasks	Segmentation, SRC writer identification online.	SRC Segmentation, Annotation	Annotation and verification, recognition and segmentation	All fields of application a SRC

Online Handwriting Recognition. In Proceedings 4th Int'l Conf. on Language Resources and Evaluation.

Biadsy, F., Saabni, R., EL-Sana, J., 2011. *Segmentation-free online Arabic handwriting recognition*. In J. Pattern Recognit. Artif. Intell,25, 1009-1033.

Cheriet, M., Thibault, R., Sabourin, R, 1994. *A multi-resolution based approach for handwriting segmentation in gray-scale images*. In Proceedings of IEEE International Conference on Image Processing.

Elanwar, R., Rashwan, A., Samia, A., 2010. *OHASD: The First On-Line Arabic Sentence Database Handwritten on Tablet PC World Academy of Science*. Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering.

Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., Janet, S., 1994. *UNIPEN project of on-line data exchange and recognition benchmarks*. In Proc. of the 12th International Conference on Pattern Recognition.

Jäger, S., Manke, S., Reichert, J., Waibel, A., 2001. *Online handwriting recognition the npen Recognizer*. In IJDAR, 3, 169-180.

Liwicki, M., Bunke, H., 2005. *Iam-ondb - an online english sentence database acquired from handwritten text on a whiteboard*. In Proceedings of the Eighth International Conference on Document Analysis and Recognition.

Marti, U. V., Bunke, H., 1999. *A full english sentence database for off-line handwriting recognition*. In Proceedings of the Fifth International Conference on Document Analysis and Recognition.

Nakkach, H., Hichri, S., Haboubi, S., Amiri, H., 2016. *A Segmentation Free Approach to Strokes Extraction from Online Isolated Arabic Handwritten Character*. In International Conference on Advanced Technologies for Signal Image Processing ATSIP.

Pechwitz, M., Maddouri, S., Maergner, V., Ellouze, N., Amiri, H., 2002. *Ifn/enit database of handwritten arabic words.* In Proceeding Of CIFED.

Plamondon, R., Srihari, S. N., 2000. *On-line and off-line handwriting recognition: A comprehensive survey*. In IEEE Transactions on Pattern Analysis and Machine Intelligence, 22.

Steinherz, T., Rivlin, E., Intrator, N., 1999. *Offline cursive script word recognition-a survey.* In Int l. J. Document Analysis and Recognition, 2, 90-110.

Tagougui, N., Kherallah, M., Alimi, A. M., 2013. *Online Arabic handwriting recognition: a survey*. In International Journal on Document Analysis and Recognition (IJDAR), 16, 209-226.

Viard-Gaudin, C., Lallican, P. M., Knerr, S., Binter, P., 1999. *The IRESTE On/Off (IRONOFF) Dual Handwriting Database*. ICDAR.

Vinciarelli, A., 2002. *Survey on off-line Cursive Word*. In Recognition Pattern Recognition, 35(7), 1433-1446.

Wilkinson, R., Geist, J., Janet, S., Grother, P., Burges, C., Creecy, C., Hammond, B., Hull, J., Larsen, N., Vogl, Y., Wilson, C., 1992. *The First Census Optical Character Recognition Systems Conference*. In the U.S. Bureau of Census and the National Institute of Standards and Technology.