

Towards a Database Oriented Research in Hadith Using Relational, Algorithmic & Data- Warehousing Techniques

By

Mr. Syed Irfan Hyder *

And

Mr. Syed Ghazanfar Ahmad *

Abstract

Extensive effort is being exerted in making available Hadith literature in a computerized form and its wide dissemination using web and other media. The effort is currently focused on enabling pre-specified indexing and searches. There is now an increasing effort to tag specific categories of words in the text with definitions and tables enabling more intelligent searches. However, these database-oriented efforts still do not expose the database design for research and exploration through established computer science technologies such as algorithmic searches and relational querying.

This paper presents the results of our research to define a graph theoretic representation of the chain of narrators of Hadiths and an aligned database structure suitable for storing

-
- * Karachi Institute of Economics & Technology, Karachi, Pakistan
 - * Karachi Institute of Economics and Technology, Shaikh Zayed Islamic Research Center, University of Karachi

the biographical data of the narrators and other historical events. Thesis of this paper is that the use of computer science concepts for algorithmic research, database queries, data-warehouses and use of advanced data-mining techniques greatly assists Hadith research and research in Islamic history and literature. It would make available our vast Hadith literature in a format amenable for cross verification and analysis in a computationally feasible manner. Certain classes of analysis that were considered intractable earlier may become feasible using current technologies.

1- Introduction

Generation of large amounts of Hadith related historical literature in the early centuries of Islamic scholarship became so voluminous that it soon became infeasible to cross tabulate and cross evaluate *all* the work done and citations on a particular person or event. Technology constraints of that time precluded the availability of authentic published material by *all* the scholars on a particular subject and its reproduction and dissemination at all the centers of learning. The demand from a scholar to possess knowledge of all the early literature also started bearing down on human limits of time and available resources. These constraints do not hold with today's technologies. Madinah has been the center of learning and can now take a leadership role in enabling the conversion of Islamic historical material to modern database technologies. Design of such databases and interfaces would enable complex research including data-warehousing enabled multi-dimensional analysis and data-mining techniques.

A Hadith consists of two main parts: the Sanad **سند** and Matn **متن**. Sanad **سند** (plural: Isnad **إسناد**) is the chain of narrators that leads to the text of the *Hadith*. The *Sanad* **سند** consists of all those who narrated the text, starting with the last narrator (who records the *Hadith* in his book, called a *muhaddith* **محدث**) and ending with the Prophet (sa). Following is the *Sanad* of a *Hadith*:

Al-Bukhari البخاري → Musaddad مسدد → Yahyaa يحيى →
Shu'bah شعبة → Qataadah قتادة → Anas أنس → Prophet
Muhammed محمد (sa).

The text of the *Hadith* or what the Prophet actually said or did is called the *Matn* متن . In the *Hadith* mentioned before, the *Matn* is "**None of you truly believes until...**"

In this paper we use Set Theory to model a chain of narrators. Recognizing that a Hadith may have multiple chains of narrators, we extend the model using Graph Theory. We propose the use of a Directed Acyclic Graph (DAG)¹ to model the relationships of multiple chains of narrators for a *Hadith*.

Given this machinery we are able to model classifications of *Hadith* referring to the evaluation of the chain of narrators. Evaluation of narrators as done in *Rijal al-Hadith* (study of the narrators of *Hadith*) requires developing a biographical database of narrators. The event model described in this database is inspired by Peter Coad's transaction model² and Len Silverston's Party Model³. The model is general enough to cover most of the biographical and historicity related events that are used in Hadith Research for evaluating people, dates, character, relationships, etc. The model is generic enough to cover roles in which a person may participate in his life.

Relational model of the *Hadith* database allows us to store all categories of *Hadith* with their multiple chains of narrators, multiple references in books of *Hadiths* and multiple evaluations by scholars. Hadith database is then linked to the Rijal رجال biographical database.

¹- R. Diestel, Graph Theory, Electronic Edition 2005, Springer-Verlag, Heidelberg, New York, 2005.

²-P. Coad, D. North and M. Mayfield, *Object Models: Strategies, Patterns, and Applications*, Prentice-Hall ECS Professional, 2nd Edition. 2000.

³- L. Silverston, W. A. Inmon, K. Graziano, *The Data Model Resource Book*, John Wiley & Sons, 1997, ISBN:0471153672

This paper describes the essential infrastructure for storing biographical and evaluation information that can be utilized by state of the art tools for analysis and research. It would enable many new kinds of analysis that were not possible before.

II- Stages in the Computerization of Manuscripts

Extensive effort is being exerted in making available the hadith literature in the computerized form. This computerization of hadith literature has primarily been focused on making available the text in computer-readable form that enables searching and cataloging. There is now an increasing effort to tag specific categories of words in the text with definitions enabling more intelligent searches. This requires imposing a structure on the documents that are initially an unstructured stream of text bytes or pixels.

The process of structuring a paper-based manuscript into computer formats typically goes through the following stages:

Making available the list, index or catalog of the manuscripts with pointer to their location and availability. This is typically the first stage. There are several such catalogs that are now available on the net. Library catalogs are a good example.

Second stage often requires scanning the document or book and saving it in a picture format. This is not a computer readable format, i.e. it cannot be sub Such quantitative grades would make it easier to compare and contrast criteria for evaluations. ected to computerized text processing such as sorting, searching, making a table of content, etc. However, this format actually saves the document and also provides limited authenticity. It allows making of copies without wear and tear of the original manuscript. Ancient manuscripts are often preserved in this format if there is a danger of their crumbling and also for researchers to easily access the material without harming them¹.

¹- www.ahlulhadith.com

Third stage is converting the scanned images through OCR or other technologies or even manually typing the entire content into a computer readable format. Computer readable format has many advantages over the scanned images. They allow textual searches, ease of copying and pasting material from one document to another, reformatting for better readability, and providing various ways of viewing and presenting the information¹. Alhumdullilah الحمد لله this stage has been satisfactorily achieved. There are several websites and sources where the computer readable format has been made available. There is also a downside to it. Because of the ease of copying, pasting and deleting text from such material without trace the authenticity of the presented material can only be verified through secondary sources and supplementary certifications from credible institutions.

Fourth stage starts tagging the computer readable text with various attributes using facilities available in word-processors. These attributes (e.g. named paragraph styles) enable the word-processor to identify and separate paragraphs of body text from headings, and to differentiate one level of heading from another. Advanced form of such tagging enables discriminating dates text from plain text, discriminating cities and names of people from plain text etc. This allows availability of several more features such as:

- Ability to automatically generate and maintain table of contents, table of figures, table of tables, index, cross-references, multi-level bullets and numberings, etc.
- Ability to move and navigate quickly among various sections, headings and other parts of the text as well as links to external documents and web references.
- Searching for user specified words or phrases or keywords.

¹- www.ihsanetwork.org

There are several dimensions of such search ability as seen from the way Google stores the information and presents to the user links to articles on various relevance criteria including user customization, reference to the context, juxtaposition of words, similar spellings, and other ways of associations. This is a branch of AI where the search tools try to learn from the search behavior of the user(s) and optimize the search based on various criteria such as frequency, importance, relevance etc. Such techniques have been perfected especially for documents that are deemed to be "unstructured" or without a formal structure. Hadith literature has been typically treated during computerization as documents with limited or no structure.

A commendable effort is the search facility in Ihsan network system [3] for names of people in *Hadiths*. The system has resolved the similarities of names and spellings and has uniquely linked a person's name to all of its references in the text. This effort represents the start of structuring process for *Hadith* literature. The system resolves all the references to a name and links them to a single definition. It is now possible for this unique definition to be linked to a wealth of biographical information stored in a relational database.

Similarly Ihsan network system¹ resolves all references to a city in the literature to a unique definition. It is now possible to link such a definition to historical and geographical information in a relational database.

Ihsan network system² also tags dates and discriminates them from the plain text allowing complex searches that can find references to dates between (say) 20 AH to 100 AH. This in turn may allow searching through either the AD calendar or AH calendar or any other.

¹- www.ihsanetwork.org Encyclopedia of hadith(soft ware)

²-Al-Nawawi, *Muqaddimah Taqreeb ur Rawi*, Beirut, Maktaba al Kauthar,2001 p.14

There is a need to further enhance the textual tagging to actually link all the words to their roots and their various forms and then linking their definitions in turn to a scholarly description of the evolution of the word and its derivatives over ages, places, people and literature.

The next stage of structuring the information is to organize the relational structure. Storage of relational structure, meta-data describing the structure and storing the information in the database is a tedious but a very important task. The result of this structuring is to enable algorithmic and special searches such as the use of SQL on databases. These searches and algorithmic processes enable not only the retrieval of information based on a given criteria but also enable prediction and identification of anomalies by cross tabulating and post processing the information retrieved. This is how transaction data is stored in business information systems enabling enterprise resource planning (ERP) and forecasting.

Next stage of research involves transferring the data from the relational database in to a data-warehouse, which can then allow multi-dimensional analysis and extraction of information using advanced data-mining techniques for extracting interesting patterns.

III- Model of Isnad (chains of narrators)

We now model a generic chain of narrators.

Let v_i be a narrator; $1 \leq i \leq n$.

Let the set of all the teachers of v_i be denoted by $v_i.T$.

Let the set of all *Sahabee* be S i.e. companions of the holy prophet.

Let the set of all *Tabiee* be B , i.e. successors of the companions of the holy prophet.

Let the set of all *Muhaddith* be M i.e. scholars of *hadith* who had written *hadith*.

The link from v_i to v_j may be defined as:

$$v_i \rightarrow v_j \Rightarrow v_j \in v_i.T \quad (1)$$

A chain of narrators (*sanad*) for a *Hadith*, then becomes:

$$v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow \dots \rightarrow v_{n-1} \rightarrow v_n.$$

Typically, v_1 is a *muhaddith* or the expert who has written down the *Hadith*, and v_n is the Prophet (sa).

The rule for v_j to be a teacher of v_i may differ according to different scholars. According to some, mere co-existence during the same era is enough to establish the link, whereas for others it must be established that not only the two co-existed, but were also reported to have met or accompanied each other. The link is strongest if v_2 has formally been a teacher of v_1 ¹.

Some traditionists prefer Saheeh al-Bukhari صحيح البخاري to Saheeh Muslim صحيح مسلم because al-Bukhari البخاري look for those narrators who had either accompanied or met each other, even if only once in their lifetime. On the other hand, Muslim would accept a reporter who is simply found to be contemporary to his immediate authority in reporting.²

Al-Hakim defines a *Musnad* or *Mutasil* (supported) *Hadith* as: "A *Hadith*, which a narrator reports from his teacher from whom he is known to have heard *Hadith* at a time of life suitable for learning, and similarly in turn for each teacher, until the *Sanad* reaches a well-known Companion, who in turn reports from the Prophet³.

A. Classifications According to the Links

Given the above definitions, a *hadith* with a chain of narrators $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow \dots \rightarrow v_{n-1} \rightarrow v_n$ can now be evaluated as per the specifications given by the scholars. Note that in the following discussion each bullet represents a condition that can be

¹- Muhammed b. Abdullah al-Hakim, *Ma'rifah 'Ulum al-Hadith*, Cairo Darul Kutub Al Ilmiya 1937 p.17

²- Ibn Hajr al-Askalani, (ed. M. Aud & M. G. Sabbagh), *Sharh Nukhbah al-Fikr*, Damascus. 1410H/1990] p.8-9

³- Al-Sunani, Toudhieh Ul Ifkar, Cairo (2 vols.), 1366 H, p. 455.

checked by the computer algorithmically given that the corresponding data and the relationships are available in the database. The database design that would provide this input through queries is described in the next section.

A Hadith is *Marfu-Mutasil* متصل معروف (Elevated-Connected) if:

- v_n is the Prophet (sa), and
- v_1 is a *muhaddith*, $v_1 \in M$, and
- Every node, v_i is authentic (*siqa*), and
- Links from v_1 to v_n follow Rule (1).

A Hadith is *Mauquf-mutasil* (Stopped-Connected) if:

- $v_n \in S$, i.e. v_n is a *sahabee* (i.e. the sanad goes up to a *Sahabee*), and
- $v_1 \in M$, i.e. v_1 is a *muhaddith*, and
- Every node, v_i is authentic (*siqa*), and
- Links from v_1 to v_n follow Rule (1).

A Hadith is *Muqtu-mutasil* (Severed-Connected), if

- $v_n \in B$, i.e. v_n is a *Tabi'ee* (i.e. *Sanad* goes up to a *Tabi'ee*), and
- $v_1 \in M$, i.e. v_1 is a *muhaddith*, and
- Every node, v_i is authentic (*siqa*), and
- Links from v_1 to v_n follow Rule (1).

A Hadith is *Mursal* (Hurried), if

- v_n is the Prophet (sa), and
- $v_1 \in M$, i.e. v_1 is a *muhaddith*, and
- Every node, v_i is authentic (*siqa*), and
- Links from v_1 to v_{n-1} follow Rule (1), and
- v_{n-1} is not a member of S (but $v_{n-1} \in B$), i.e. v_n is not a member of $v_{n-1}.T$. That is, the link of *Sahabee* between the *Tabi'ee* and the Prophet (sa) is missing, e.g. when a *Tabi'ee* says, "The Prophet said....".

A Hadith is Munqata حديث المنقطع (broken), if

- v_n is the Prophet (sa), and
- $v_1 \in M$, i.e. v_1 is a *muhaddith*, and
- Every node, v_i is authentic (*siqa*), and
- Links from v_2 to v_n follow Rule (1), and
- A link closer to Muhaddith, e.g. $v_1 \rightarrow v_2$ does not follow Rule (1) i.e. v_2 is not a member of $v_1.T$.

A Hadith is Mu'dal الحديث المعضل, if

- v_n is the Prophet (sa),
- $v_1 \in M$, i.e. v_1 is a *muhaddith*
- There is some j : $1 < j < n-1$, such that two consecutive links are missing after j . That is v_{j+1} is not a member of $v_j.T$, and there exists some unknown narrators v_k, v_t not in chain that could have made the following hold under Rule 1: $v_j \rightarrow v_k \rightarrow v_t \rightarrow v_{j+1}$.
- The number of consecutive missing narrators in a *Sanad* exceeds one (typically two).

A Hadith is Mu'allaq الحديث المعلق (hanging), if

- v_n is the Prophet (sa),
- $v_1 \in M$, i.e. v_1 is a *muhaddith*
- $n = 2$. That is, the *Muhaddith* omits the whole *Sanad* and quotes the Prophet (sa) directly.
- Also, called *balaghah* (to reach); e.g. **Imaam** Maalik sometimes says in al-Muwatta', "It has reached me that the Messenger of Allaah said".

B. Classifications for Multiple Chains of Narrators

There may be multiple chains of narrators for a given Hadith. Classifications such as Mutawatir متواتر and Ahad آحاد need to test the number of narrators at various stages of these multiple chains. We, therefore extend our Set Theoretic model

of a chain of narrators to a Graph Theoretic model that can represent multiple chains of narrators. We use a Directed Acyclic Graph (DAG) to represent the multiple chains. A directed path of this DAG then becomes our previously defined chain of narrators.

Let G be a DAG representing multiple chains of narrators for a Hadith. G consists of a set of vertices (V), which that represent the set of narrators, and a set of arcs, where each arc represents the relationship given in Rule (1).

We note that G is actually a *rooted* DAG. The *Muhaddith* of a *Hadith* may be considered the root of its DAG and is said to be at level 1. The root node is now labeled in a special way so as to distinguish it from all other vertices in V . Vertices (narrators) directly connected to the root (i.e. *Muhaddith*) are considered to be at level 2. Vertices.

Directly connected to those at level 2 are considered to

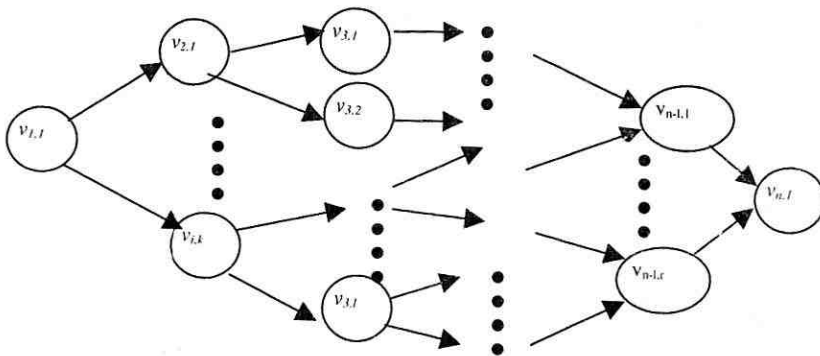


Fig. 1. Multiple chains of narrators form a Directed Acyclic Graph

be at level 3, and so on.

We denote the *Muhaddith* (root) at level 1 as $v_{1,1}$. Let the number of levels of G be n . Narrators at level $i: 1 \leq i \leq n$ are denoted by $v_{i,p}$ where $1 \leq p \leq m_i$ (m_i being the number of vertices at level k). Fig. 3 shows the DAG representing multiple chains.

Given that $v_{i,p}, v_{j,q} \in V$, The arc $(v_{i,p}, v_{j,q})$ is an ordered pair where $v_{i,p} \rightarrow v_{j,q}$ follows Rule (1). A chain c_i narrators $v_{1,1}$

$\rightarrow V_{2,3} \rightarrow V_{3,4} \rightarrow \dots \rightarrow V_{n-1,2} \rightarrow V_{n,1}$ is a path of this graph.

Given the above definitions, we can now model various classifications of Hadiths according to the number of narrators as follows:

A *Hadith* is *Mutawatir* (consecutive), if for each level k , the number of narrators is greater than a minimum number " m " specified by an established Hadith scholar. In Hadith literature four, five, seven, ten, twelve, forty, or even seventy has been suggested as a minimum [9].

A *Hadith* is *Ahad* or a *Khabr Waahid* if for each level k , the number of narrators is less than " m ", the minimum specified limit for a *Mutawatir* hadith.

Ahad is further classified into *Gharib* (strange), *Aziz* (rare), and *Mashhur* (famous).

A *Hadith* is *Gharib* (scarce, strange), if there exists some level k where the number of narrators is only one.

A *Hadith* is *Aziz* (rare, strong), if there exist some level k , where the number of narrators is two [10].

A *Hadith* is *Mashhur* (famous), if there exists some level k , where the number of narrators is greater than two

Note that a *hadith* may start out as *Aziz* initially but may cross over to become *Mashhur* at a later level ¹.

IV. RELATIONAL DATABASE MODEL

The idea is to capture the life history or biography of each narrator and those related to *Hadith* literature in a relational database shown in Fig.2 which allows researchers to use SQL queries for verifying various conditions on which *Hadiths* and specifically their chains of narrators are evaluated.

We use a simplified version of the party model [2] to flexibly cover all types of roles of a person or an organization,

¹ Al-Iraqi, A; *Taqeed wal Izah*, Darul Hadith, Beirut, 1984, p.229.

whose generalization is the party entity. A person plays many roles in his/her life. A person participates in an event in a specific role. There are a number of participants in an event. Role types include father, mother, subject, executioner, author, scribe, narrator, evaluator, teacher, *sahabee*, *tabi'ee*, *tabatabi'ee*, *muhadith*, student, judge, courtier, warrior, writer, reader, etc.

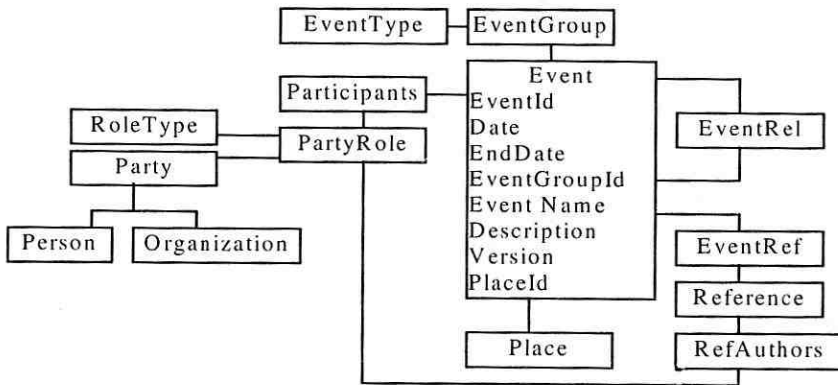


Fig. 2: Structure of Biographical Database.

Our event model in the biographical database of *rijal-al-Hadith* رجال الحديث is inspired by Peter Coad's transaction model [1]. An event is an occurrence that happens at a certain time and place and is worth remembering and is considered of some value. Events can be of many types such as birth, childrearing, mentoring, death, meeting, schooling, teaching, writing, commenting, war, fighting, visiting, etc. Events can also be related to *Hadith* research such as narrator evaluation, *hadith* narration, *Hadith* evaluation. As we can see events can be defined very flexibly to cover a large number of categories such as biographical, historical, and those related with *Hadith* narration and evaluation.

The model allows us to define a large number of events at a granularity necessary for the system to automatically check conditions such as whether two people were "*Muasiir*" i.e. co-existing

during the same period, or co-located during the same time and place, or whether a person studied from the other or not.

An event has multiple participants. For example, the birth event may have participants such as the person himself (in the subject role), father, mother, and optionally the midwife (if any).

Therefore, to find whether x and y coexisted during the same period, we can find the birth date and the death date from the corresponding event types for x and y and then do comparison among the values of birth.event.date, death.event.date of x and y where they were the in subject roles. To find out where x has studied, we just need to query x 's role as student in the database to find out all events where x participated as a student. Similarly, to find out all instances where x participated as a "*Hadith* narrator".

Note that there can be multiple references of an event in different books. Events are also related to each other. An event can be related to several successors, and/or predecessors indicated by *Ventral*.

A reference book may have multiple authors, where authorship is itself a role, enabling the biographical history of the author to be also stored in the event database.

Hadiths, their chain(s) of narrators, along with their evaluations is modeled in Fig. 3. A Hadith is referred in multiple books of Hadith. Hadith, its *Matn*, narrators and each link is separately evaluated by scholars. These evaluations with their references is stored in the relevant tables. Note that there can be multiple evaluations for each one of the entities that describe a Hadith. Note that evaluations in *Narrator Val*, *Manual*, *HadithEval*, *LinkEval* and *ChainEval* contain judgments, whereas actual event of evaluation with the relevant participants, references is captured in the event table biographical database.

This unique definition of a narrator in turn can now be linked to a wealth of biographical information stored in a

relational database and linked to all the sources where there are references to the person as well as mentions about him and his life and works as indicated by the link of narrator to Party Role in Fig. 2.

V. Classification according to Evaluation of Narrators

Evaluation of a *Hadith* depends on the evaluations of its narrators. Verdicts such as *Saheeh* (sound), *Hasan* (good), *Da'if* (weak) and *Maudu'* (fabricated) rest mainly upon the nature of the reporters in the *isnad*. Evaluations of Hadith are entered in HadithEval table and evaluations of narrators in the NarratorEval table.

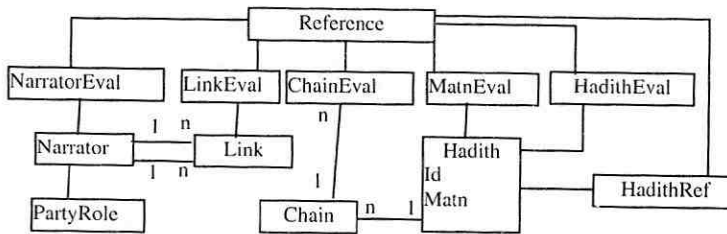


Fig. 3: Structure of Hadith Database.

Rijal al-Hadith (study of the narrators of *Hadith*) has resulted in authenticating or disparaging evaluations of narrators by different scholars. These range from

- (i) Imaam, leader,
- (ii) Hafiz, preserver,
- (iii) Reliable, Trustworthy,
- (iv) Makes mistakes,
- (v) Weak,
- (vi) Abandoned (by the Muhaddithoon), to (vii) Liar (used to fabrication). Circumstances leading to such evaluations are stored in the event table (Fig. 2) of our database and the resultant judgments in the NarratorEval table (Fig. 3).

Queries in to the tables could pull out these evaluations for a narrator.

If all the evaluations of a narrator by established scholars pulled out from the NarratorEval table are unanimously either Imam or Hafiz, then this may contribute to a *saheeh* (sound) chain provided all narrators meet the criteria. If a chain contains a narrator whose evaluation is either abandoned or liar, the chain is likely to be *da'if jiddan* **ضعيف جدا** (very weak) or *maudu'* (fabricated). If a narrator is evaluated to be one who makes mistakes, then this may cause the chain to become *da'if* **ضعيف** .

However, several such narrators relating the same *Hadith* independently may raise the rank of the Hadith to that of Hasan (**حسن** good). The database allows an algorithm to go over the evaluation of each node and arc of the graph and annotate them with ranks and aggregate an average rank giving the overall grade. Such aggregation provides useful comparison among the various opinions.

Strength of a Saheeh Hadith **حديث صحيح** varies from 1 to 7¹ as given below. We see from the list that the strength criteria are easily evaluated through SQL queries on the tables of our database:

1. Referenced of both al-Bukhari and Muslim;
2. Referenced by al-Bukhari only;
3. Referenced by Muslim only;
4. Not referenced by al-Bukhari and Muslim but fulfilling the requirements of both;
5. Only fulfilling the requirements of al-Bukhari;
6. Only fulfilling the requirements of Muslim;
7. Declared *Saheeh* by other established *traditionists*.

¹- Al-Tibi, al-Husain b. 'Abdullah, al-Khulasah f1 usul al-Hadith (ed. Subhi al-Samaira'I), Baghdad, 1391 H, p. 36)

Fig. 3 shows that the database allows a Hadith to be mentioned in more than one reference book. Al-Bukhari's requirements for links and chains is evaluated using the machinery given in the Section III, while the information required for testing each condition on the narrator, chain, link or *Matn* is pulled from the Hadith database (Fig. 3) or Rijal Biographical Database (Fig. 2).

A *Hadith* is *Hasan* حديث حسن, if:

- It has not been evaluated to be *shadhdh*, and its sanad does not contain a narrator evaluated to be disparaged, and there are more than one chain, or
- Its chain of narrators is neither *mursal* nor *munqati'* nor containing a *mudallis* [13], or
- It is evaluated to be free of *shudhudh* and it has a sanad containing a narrator evaluated as *mastur* ("screened", i.e. no prominent person reported from him), and evaluated as not totally careless in reporting, and there is another corroborating *sanad* [13], or
- It is evaluated to be free of *shudhudh* and has a *sanad* containing a narrator evaluated to be truthful and reliable, but is a degree less in his preservation/memory of *Hadith* in comparison to the reporters of *Saheeh Hadith*¹.

Hadith Eval would provide judgment about freedom from *shudhudh*, evaluations of chains has already been discussed, *mastur*, carelessness in reporting, truthfulness, reliability are evaluations of the narrators. We can therefore verify each of these conditions by querying the database. To enable judgment regarding *Hasan*, these conditions would have to be tested together that would require developing an algorithm to test the given criteria.

Narrator Val table allows storage of different evaluations

¹- Jalal ad-Din al-Suyuti, Tadrīb al-Raawī, Cairo, 1379H/1959, 1: 171.

of the same narrator by different scholars.

A *Hadith* is *Da'if* ضعيف if it fails to reach the status of *Hasan*. Or if it is either of type *mursal* مرسل , *mu'allaq* معلق , *mudallas* مدلس , *munqati'* منقطع or *mu'dal* معضل . Or a narrator has a disparaged character because he has been evaluated to be a liar, or fabricator, or makes mistakes, or has opposition to the narration of more reliable sources, or there is ambiguity surrounding his person.

Smaller the number and importance of defects, the less severe the weakness. The more the defects in number and severity, the closer the *Hadith* will be to being *maudu'* (fabricated)¹. Degree of such weakness requires that we assign a rank or grade to each defect, then an aggregate average of the rank can be obtained. This may be weighted-average if different criteria are considered to be of different importance.

A *Hadith* is *maudu'* حديث موضوع if:

- *Matn* contradicts the established norms of Prophet's sayings, or its narrators include a liar.
- Its fabricator has confessed.
- There is discrepancy of dates or times of incidents².

Matn evaluation would contain information about the normative nature, narrator's evaluation would in *NarratorEval*, fabricator's confession is captured in the *Biographical* database where the role of that person is fabricator, and his saying is an event. Discrepancy of dates and times are given in *Matn* evaluation, while discrepancy in the *biographical* dates and times of events can be recognized from the *event* database.

Collections of fabricated *Hadith* (e.g. Ibn al-Jauzi in *al-Maudu'at*) can also be entered in the database along with their evaluations.

VI. Conclusions and Future Work

Thesis of this paper is that the use of computer science concepts for algorithmic research, database queries, data-

¹ - Al-Sakhawi, Fathul Mughis, Beirut, Muktaba-e-Al-Sunnah, 1995, Vol. 1: p.106.

² - Ibn Salah, Uloom ul Hadith, Beirut, p.89.

warehouses and use of advanced data-mining techniques greatly assists Hadith research and research in Islamic history and literature.

The infrastructure for this research requires a graph theoretic modeling of the chains of narrators and relational modeling of all the relevant entities in the biographical database of the Rijal-al-Hadith رجال الحديث and a Hadith database with corresponding links to references in books of Hadiths by established scholars.

Issues that need to be tackled include need for massive data entry followed by reconciliation of the entered data with authentic sources.

Formalization of various definitions in a format amenable to computerized processing also represents a huge task especially, while the data is being entered, cleaned and while the testing is being done to refine the algorithms. Initially, the results may be surprising, or even alarming, because the data, algorithms and the data structures are being refined.

Once we have achieved a reasonable level of confidence in the integrity of the data and the algorithms, we would be in a position to carry out analysis that could not be done earlier because of the lack of tools.

Possible analysis includes analysis of the paths and sub-paths within the DAG of narrators for a given Hadith. Aggregate statistics about the occurrences of such paths and their contributions to certain types of evaluations.

Annotation of the nodes and arcs with various kinds of weights and then evaluating the aggregate averages over different paths and over the entire graph to yield numerical grades of evaluations. Currently the classifications of *Hadith* are qualitative, these kinds of aggregate functions would enable quantitative grading of these classifications.

Madinah has been the center of learning and early Hadith scholarship pioneered the meticulousness of historical reporting. Conversion of Islamic historical material to modern database technologies at the granularity of the proposed event model represents another such pioneering effort.