人文学にとっての「リンク」の意義
SAT 大蔵経データベースを手がかりとして

永崎研宣
一般財団法人人文情報学研究所

Paul Hackett
コロンビア大学

苫米地等流
一般財団法人人文情報学研究所

A. Charles Muller
東京大学大学院人文社会系研究科

下田正弘
東京大学大学院人文社会系研究科

Web 上には様々な文化資料が蓄積されつつあるが、それらをどのようにリンクしていくことで有益なものとなるかという観点からの議論はまだ十分に行なわれているとは言えない。ここでは、人文学の立場からのリンクの意義について検討し、資料批判の観点を反映する形でリンクの性質を適切に記述し表現することの必要性を論じた。

The Significance of “Linking” in the Humanities

Kiyonori Nagasaki
International Institute
for Digital Humanities

Paul Hackett
Columbia Uniersity

Toru Tomabechi
International Institute
for Digital Humanities

A. Charles Muller
Graduate School of
Humanities and Sociology
University of Tokyo

Masahiro Shimoda
Graduate School of
Humanities and Sociology
University of Tokyo

While huge cultural resources have been accumulated on the Web, it does not suffice to merely discuss how they should be linked each other. In this paper, we treat significance of linking of the resources from the viewpoint of the humanities in the case of the SAT DB and conclude necessity of description of characteristics of links of the resources reflecting the notions and methods of textual criticism.

１．はじめに

近年、Web 上には膨大な文化資料が集積されつつある一方、LOD に象徴されるように、リンクが再び脚光を浴びるようになっている。しかし、テクニカルな方面や制度面からの旗振りは盛んになってきているものの、それが人文学にとってどのような位置づけとなり得るかということについては、事例の蓄積が少なく、議論もまだ十分にはされていないように思われる。そこで、本稿では、Web 上での資料間・データベース間に静的・動的に張られる関係のすべてを「リンク」と呼ぶことにした上で、人文学にとっての「リンク」がどのような意義と課題を持ち得るかについて、SAT DB（SAT 大蔵経テキストデータベース）での事例を手がかりとして検討したい。

２．SAT DB とは

SAT DB は、大正末期から昭和初期にかけて

編纂された漢文仏教経典の校訂テクストの集成である大正蔵（大正新脩大藏經）を、デジタル翻刻したものである。1 巻約 1000 頁のものが 85 巻、約 1 億字のうちほとんどが漢字、一部に悉曇文字やカナ等が用いられている。このすべてが 1994 年から 2007 年にかけてテクストとして入力され、現在は Web サービスとしても提供されるに至っている[1]。

この SAT DB は、時間をかけて丁寧に構築されたテクスト DB であると同時に、様々なデータベースとのリンクを通じて利用者に大きな利便性を提供する統合型 Web サービスという側面も持っている。本稿が目標とするのは、この SAT DB の Web サービスにおけるリンク、特に、本年 9 月に新たに試験サービスを開始したチベット大蔵経とのリンク及び他の画像データベースとのリンクに着目して検討することだが、その前に、まずは SAT DB がどのようなサービスを統合しているかについて簡単に見てみよう。

3．SAT DB におけるリンク

サービスとしての SAT DB では、様々な形で他の Web データベースとのリンクを提供することで利用者の利便性を高めている。以下に、そのリンクの種類を起点ごとに分けてリストしてみよう。

a. テクスト本文中の任意のテクストから：
1. DDB(電子仏教辞典)[1]
2. BDK-SAT Parallel Corpus
3. INBUDS（印度学仏教学論文 DB）[2]
4. CiNii[3]
5. SARDS3[4]
6. INBUDS キーワード共起情報

b. テクスト本文中の任意の文字から：
1. CHISE[5]
2. HMS (Han Morphism System)[6]
3. Unihan Database[7]
4. HNG (Hanzi Normative Glyphs)[8]

c. 各経典の書誌情報から：
1. BCRD (Buddhist Canons Research DB)[9]
2. 近代デジタルライブラリー[10]・縮刷大藏經
3. 各種画像データベース[11]

このようにしてみてみると、SAT DB からのリンクの起点になる情報には大別して 3 種類あることになる。まず、b. の任意の文字からのリンクに関しては、基本的には文字同士での対応となる。テクスト翻刻をする時点での文字に関する判断のポリシーや、文字コードの策定の段階での問題は様々に生じ得るが、異体字同時検索をはじめとする問題回避のための技術も現在では様々に用意されている。したがって、すでに出来上がったデジタルテクストと、文字をインターフェイスとするデータベースとの間のリンクと考えた場合には、双方ともに問題は少なく、比較的容易にリンクを形成して有益な情報を提供できると言えるだろう。

一方、a. の場合には、リンク先のデータについては、辞書の検索結果、パラレルコーパスの検索結果、書誌情報 DB の検索結果等がある。いずれも、主に漢文の本文の一部をドラッグした場合に、その選択されたテクストについての問い合わせが各データベースに行なわれ、その結果が適宜表示される。この場合、漢字での異体字がどう扱われているのかといった文字統合ポリシーがデー

---

1 Muller により作成・運用されている大規模佛教用語電子辞典。http://www.buddhism-dict.net/ddb/

2 詳しくは後述。http://www.inbuds.net/

3 国立情報学研究所論文検索サービス http://ci.nii.ac.jp/

4 ハレ大学が中心となってサービス提供している西欧語の南アジア研究論文書誌データベース。6 万件超の書誌情報が蓄積され検索可能となっている。http://www.sards.uni-halle.de/sards/

5 守岡知彦氏により開発・運営されている大規模文字オントロジー。http://chise.zinbun.kyoto-u.ac.jp/

6 白須裕之氏により開発された漢字データベース http://www2.dhii.jp:3000/

7 Unicode コンソーシアムにおいて開発公開されている漢字データベース http://www.unicode.org/charts/unihan.html

8 北海道大学をはじめとする国内各地の研究者により開発運営されている通時的漢字字体データベース http://www.joao-roiz.jp/HNG12/search/start

9 詳しくは後述。http://aibs.columbia.edu/databases/New/

10 国立国会図書館が開発公開している明治大正期の図書を主な対象とした大規模デジタルライブラリーhttp://kindai.ndl.go.jp/

11 これらについては後述する。

タベース毎に異なっている場合があるため、注意が必要となる。データベースによっては異体字同時検索機能を実装しているものもあり、そういったデータベースではリンクに際して特に大きな問題は生じていない。

a.6 の INBUDS キーワード共起情報は本年 9 月に搭載した試験サービスだが、これは異体字同時検索機能を有した Web API として提供されている機能であり、これもリンクとしては問題なく利用可能である。ただし、この a.6 において特に留意すべき問題点として、リンク先のデータの意味するところや本文との関係が、それを提供している文脈なしには理解が難しいという点がある。a.6 についてもう少し詳しく解説しておくと、まだ試用版だが、これは日本印度学佛教学会が提供する収録件数 7 万件を超えるインド学仏教学分野論文書誌データベースである INBUDS の Web API を利用している。INBUDS では、1 論文あたり最大 20 件程度のキーワードが付されている。ここで、1 論文毎の各キーワードの共起頻度を計測して蓄積し、検索要求に対して共起頻度 10 位までのキーワードを返戻するようになっている[1]。INBUDS 上ではこの機能を用いて検索結果表示時に関連しそうな語を提示することで利便性の向上を図っている。一つの分野における 7 万件の書誌情報に対するキーワード情報から生成されたこともあり、この関連語句は、上位 10 位までならばそれなりに妥当性があると評されている。この機能が Web API としても利用できるのである。したがって、SAT DB 上での a.6 の機能としては、本文中の単語をドラッグした際に、この Web API を利用して検索語との共起頻度の高いキーワード（≒関連度の高い術語）をリストして関連する術語を検索候補としてポップアップウインドウ中に表示するということになる。それらの語をクリックすることによる本文検索も可能であり、特に初学者には好評な機能である。

しかし、そもそも「関連の深いキーワード」についての事前に十分な説明がないことには、あらぬ誤解を招く可能性がある。さらに、INBUDS の側としては試供版として提供している機能であるに過ぎず、現時点ではキーワードの正規化ができていないため、必ずしも適切な情報を提供できているわけではない。このことについての留保も必要となるだろう。したがって、「リンク」に際しては、本来、こういった前提となる情報を認知されやすい形で提供することが必要だろう。

次に c. の「各経典の書誌情報から」について検討しよう。SAT DB の元となった大正蔵は、既存の漢訳大蔵経を当時可能な範囲で最大限収集して校訂を行なったということで、漢訳大蔵経の決定版として世界中で利用され続けてきており、その経典の番号も世界中で共有されている。それ故、チベット大蔵経やパーリ聖典と漢文大藏經との間の「実質的に同じ内容の経典」の対応づけ情報が、大正蔵の番号に依拠して対照目録として作製されてきたという経緯がある。つまり、SAT DB と他の経典とのリンクに関しては、紙媒体時代の研究の蓄積をそのままデジタル化するだけでかなりの部分が出来上がってしまうのである。本稿でそのリンク対象の事例として採り上げるのは、本年 8 月に公開された新版 BCRD である。

4．BRCD とのリンク

BCRD はチベット大蔵経に関する唯一の総合的なデジタル目録であり、一次資料と二次資料の詳細情報とを横断的にリンクするだけでなく、オンライン資料へのハイパーリンクをも含んでいる。標準的な目録システムは「アイテム」（モノグラフ、叢書等）レベルのみでの書誌情報を提供する。このレベルの目録情報では、仏典等の古典には不十分である。たとえば、チベット大蔵経は、様々な著作者の手になる約 5000 の個別の作品から成り立っている。ある仏典の校訂テクスト叢書が標準的な図書館目録では一つか二つの書誌記録として表現されるようなことは珍しくはない。個々の作品にアクセスするための関連するあらゆる詳細情報は、二次的な参照文献に属するものとなっている。同じ事は、テーマによって、あるいは著者によって分類される巨大なコレクションにおいてしばしば刊行されてきた、後代の大量の注釈文献のほとんどにおいても言える。

BCRD はチベット大蔵経に関する完全な記述だけでなく、同様に、3500 に及ぶ後代の作品とのクロスリンクをも集積してきた。そしてまた、既存の研究を十分に記述することによって発展的な研究を可能にするために生の書誌情報を提供している。そして、多言語での印刷物やデジタル資料についての書誌的な情報を継続的に更新していくというサービスを提供している。そのようなプロジェクトは、詳細で正確な、あるいは、迅速で簡潔な参照情報を提

これを簡便に実装した上で Web API として活用できるようにしたものであると言える。

[1] INBUDS のキーワードの情報を活用したサービスに関しては、相場徹氏らによる先行事例があった[4]。今回の新たなサービスでは、

供することで人文学研究者の幅広い領域を益する。

加えて、BCRD は 1500 万シラブルに及ぶチベット大蔵経全体の全文テクスト検索を提供している。現在のところ、700 人以上のアクティブユーザがおり、最新の自然言語処理の技術が提供されている。これは知的な検索や検索結果を的確に捉えることを可能にするためである。結果として、ユーザからのフィードバックはきわめて肯定的であり、あるユーザは「革命的」であり、「大変に先進的で、この分野の研究者に多大な利益をもたらしている」資料であると評している。

さらに、BCRD についてもう少し具体的に説明すると、たとえば、漢字で「中論」と検索すると Nāgārjuna の『中論』がヒットし、このテクストのサンスクリット語やチベット語での名称が表示されるとともに、チベット語訳のテキストデータベース（主に ACIP）や画像データベース（主に TBRC）へのリンクと中国語訳のデータベース（SAT 及び CBETA）へのリンクも表示される。さらに、『中論』に対して書かれた後代の注釈書のタイトルもリストアップされ、それらについてもまた、『中論』に対するものと同様の関連情報が提供されるようになっている。そして、まだ十分には機能していないが、関連する二次文献の情報もリストされるようになっている。Web 上の様々なリソースの多くがこのようにして統合的に活用できるようになっていることは、利用者にとっては大変望ましい状況であると言える。

この BCRD と SAT DB が、今回、既存の対照目録[2]をベースとして作製されたデジタル対照目録を通じて経典毎に相互リンクされたというのが今回報告する新規機能である。具体的には、ユーザがいずれかの経典を SAT DB 上で表示させると、図 1 に掲示したポップアップウインドウ上に、その経典に対応する BCRD 上の経典情報へのリンクが表示されるようになっている。

このことは、漢文の大蔵経を読みながら、必要に応じて容易にチベット語訳のテクストも参照できるということを意味しており、そして逆もまた然り、である。これは一見すると単にリンクを張っただけであるように思えるが、実際のところ、

図 1. 他の DB への経典毎のリンク用ポップアップウインドウ
Figure 1. A pop-up window for listing links to other databases

デジタル版がなかった時代には、それぞれ膨大な分量の両大蔵経をいつでもアクセスできるようにした上で対照目録を参照しながら対応する経典を探して閲覧するという作業はそれだけで大変なものであった。デジタル大蔵経が完備された後であっても、今回の機能が提供されるまでは、一々対照目録を参照して経典番号を確認してからデジタル大蔵経の当該経典を検索して閲覧するという作業が必要であり、これも少なからぬ手間がかかっていた。この一連の作業が数回のクリックのみで実現できるようになったことは、研究環境の大きな変化であると言える。さらに言えば、すでに安定したサービスを広く提供し、ユーザを多く抱えている漢文とチベット語の大蔵経データベースに対して提供された新たな機能ということであり、これまでのデジタル大蔵経閲覧の延長として操作することができるのは、ユーザにとっては新しいシステムの使い方を習得したりする必要がなく、その意味では利用者コミュニティ全体にとっての負荷があまり増えないということも特筆すべき点だろう。

しかし一方で、「このチベット語訳とこの中国語訳は同じ経典から翻訳されたものである」という判定がどのように可能なのか、そしてその判定にはどれくらい信頼がおけるのか、ということは、その都度その都度検討を重ねていく必要がある。というのは、研究の進展につれて、様々な理由により、既存の対応判定では必ずしも適切でない場合が出てきており、また、学説によって判定が割れる場合もある。BCRD と SAT DB がリンクされたことにより、そうした基盤的な研究が一層刺激されるのではないかということも、十分に期待できるところである。

それでは、Web 上での「リンク」は、そのような事情を適切に表現できるのか。まず、この「リンク」そのものに関して、その根拠としての出典や判定者といった情報が必要になるだろう。近年の学術論文やその著者としての研究者に関しては、DOI や ORCID によって一意に参照できるようになってきていることから、根拠情報そのものについては解決の糸口が見えていると言っていいだろう。さらに、どれくらい確実性があるか、ということについても記述できた方がよい場合もあるかもしれないが、そういった事柄を記述する場合には、TEI P5 (Text Encoding Initiative P5 Guidelines) の「第 21 章 Certainty, Precision, and Responsibility」[1]において整理されている人文学資料の曖昧さの記述方法が一つの大きな手がかりとなるだろう。ただ、それをどのようにして Web データベース上に見やすくわかりやすく表示するか、というインターフェイス面については今後の課題としておきたい。グラフのような形で表示して、ノード間の記述の仕方を工夫するというのが一つの方法であると考えられる。D3.js がこの種の事柄に役立ちそうなさまざまなプラグインを提供しているので、現在はそれらの活用を検討しているところである。

5．国立国会図書館デジタル資料とのリンク

2013 年 2 月 21 日、国立国会図書館（NDL）近代デジタルライブラリーにおいて、著作権処理を完了した書籍約 2.3 万点が公開された。この中には、仏教研究において直接有益なものとなる重要な図書が数多く含まれていた。なかでも筆者等が特に注目したのは『大日本縮刷校訂大蔵経』(縮刷蔵, 全 419 巻, 1881～1885 年刊行) であった。

初版の縮刷蔵は、それを拡大印刷した頻伽精舎刊の大蔵経が大正蔵の原稿として用いられ、また、その頭注に記載された、増上寺所蔵の高麗版大蔵経、宋・思渓版大蔵経、元・普寧版大蔵経との校勘情報の多くは大正蔵に継承されたということもあり、大正蔵のテクストや活字字形の来歴を検討する上で欠かせない資料となっている。しかし、NDL がデジタル化公開した縮刷蔵は、そもそも全巻揃いではなく、216 巻分しかない。さらに、大正蔵の原稿の元となった初版本ではなく 1939 年に再版されたものであり、初版本に比べて色々な訂正が施されていることから、縮刷蔵のテクストとしての正確性は高まったものの、大正蔵の来歴を確認するための比較対照という観点からはやや不十分なものとなっている。加えて、近代デジタルライブラリー版の撮影画質が文字の字形の微細な違いを確認するには、若干足りないということもあり、本来期待し得る条件を十全に満たしているとは言えない状況である。

以上のことを踏まえた上でも、近代デジタルライブラリーのデジタル化縮刷蔵は、我国の活版印刷の初期における大規模出版物の希有な事例の一つとして、そしてまた、大蔵経のテクストの問題を解決する上での手がかりとして、十分な有用性を持っている。そこで、このたびの SAT DB への試験的機能付加の際には、BCRD とのリンクに加えて、この近代デジタルライブラリーの縮刷蔵への経典ごとのリンクをも追加したのである。

このリンクもまた BCRD と同様に、SAT DB のユーザインターフェイスに組み込む形で追加した。これも図 1 に表示されているので参照されたい。縮刷蔵は、他の多くの大蔵経と同様に、本

[1] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CE.html

の単位と内容となるテクストの単位とが一致しておらず、一冊の中に数本の短いテクストが収録されていたり、数冊〜数十冊を使って一つの長いテクストを収録したりしている。元々、近代デジタルライブラリーでは、多くの図書で目次情報を記入して Web API から取得できるようにしており、また、一つの図書の中の特定の画像に対して URL でリンクを張れるようになっていることから、多くの図書では、各章の見出しを取得して、その開始頁を含む画像に対して自動的に外部からリンクを張ることができるようになっている[1]。しかし、この縮刷蔵の場合には、現時点では、中に含まれている経典のタイトルやその開始頁へのリンクといった情報が含まれていない。そこで、筆頭著者が実際にそれを確認して対応リンク情報を集積していくという作業を行なった。

なお、これに関しても、紙媒体による対照目録が出版されており[3]、これまでもそれを参照することで対応するテクストを探し出して閲覧することは可能であった。ただし、記載されていたのは巻番号までであり、巻の中の何頁からそのテクストが始まっているのかという情報は対照目録には記載されていない。一方、ここで目指す利便性としては、対応テクストの開始頁へのリンクを基本としたため、結局の所、対照目録は使用せず、具体的に近代デジタルライブラリーの縮刷蔵そのものを確認しながらの作業となった。

この場合も、本来は縮刷蔵と大正蔵を所蔵している場所で対照目録と突合わせながら閲覧をしたり、あるいは、近代デジタルライブラリーのビューワの頁をめくっていったりすることを考えたなら、かなりの労力の節約となるだろう。これを通じて縮刷蔵の意義の再評価につながったなら、これもまたデジタル化の恩恵と言えるだろう。

６．各地の画像データベースへのリンク

さらに、「リンク」を考える上で採り上げたい SAT DB のもう一つの新機能について説明しておこう。これは、各地の画像データベース[2]において公開されている仏典関連の画像資料を探索し、そのパーマネント URL、もしくはそれに類するものを集め、SAT DB の各経典の頁にて、同じ経典の画像資料の URL があれば上述のポップアップウインドウ上に表示させ、それらに対してダイレクトにリンクが張られるようにしたというものである。これによって、各地の仏典の写本や木版の画像を容易に確認できるようになった。しかし、ここでのリンクの性質には上記のリンク以上に様々な種類があり、ただ並置するだけでは、ユーザ側にかなり幅広い知識を要求することになってしまう。典型的な例を３つ挙げてみると、大正蔵第 85 巻所収の敦煌文書翻刻テクストの一部は、対象となった写本画像が大英図書館やフランス国立図書館の画像データベースで公開されている。また、大正蔵が参照した高麗版大蔵経の異本も数点が対象になっている。あるいは、大正蔵にはほとんど反映されていない江戸時代の版本の画像についても対象としている。もちろん、すでに研究成果が蓄積されているものが少なくないため、今後は、それらを適宜反映する形でリンクを記述していくことを目指したい。

終わりに

以上のように、仏教学にとってもリンクは様々な利点があるが、デジタル媒体の特性を活かした仏典テクストの提供には、同じテクストを表現するものとして括られるテクスト群・画像資料群のリンクを適切に記述することが重要となる。このことは総じて資料批判の手続きを可視化する途中経過であると言うことができ、その意味では、仏典に限らず、人文学の他の様々な分野においても同様に適用することが可能だろう。今後は、TEI P5 第 21 章をはじめ、様々な先行事例を参照しつつ適切な記述方法を検討し、SAT DB においてその成果を実装していきたい。

参考文献

1) Nagasaki, Kiyonori, Toru Tomabechi, and Masahiro Shimoda. “Towards a Digital Research Environment for Buddhist Studies.” Literary and Linguistic Computing 28.2 (2013): 296–300. llc.oxfordjournals.org. Web. 9 Feb. 2014.
2) 東北帝国大学法文学部, ed. 西蔵大蔵経総目録. 仙台: 1934. Print.
3) 高楠順次郎, and 渡辺海旭, eds. 昭和法宝総目録. 大正新脩大藏經刊行会, 1929. Print.
4) 相場徹, and 生出恭治. “インド学仏教学論文データベース INBUDS を用いた術語間関係の大きさの推定について.” 情報処理学会研究報告. 人文科学とコンピュータ研究会報告 98.11 (1998): 7-14. Print.

[1] 筆頭著者が開発した「翻デジ 2014」では、この機能を活用して文書の構造化の一部を自動的に実行できるようになっている。

[2] 現在対象となっている画像データベースは、以下の各機関が公開しているものでり、総計約 600 点である。（NDL、大英図書館、フランス国立図書館、HathiTrust、国文学研究資料館、龍谷大学、早稲田大学、立命館大学、京都府立総合資料館、学習院大学）