Vol.2013-CH-97 No.1
2013/1/25

リソース連携を通じたテキスト・データベースの新たな可能性に向けて ーSAT2012 を事例としてー

永崎研宣† 苫米地等流† 下田正弘††

SAT2012 は，テキスト・データベースの利便性を高めるべく，パラレルコーパス等を構築するとともに，様々な他のリソースとの連携機能を盛り込むなどして公開された新たな Web サービスである．本稿では，SAT2012 構築に際して採用した人文系データベースの構築手法や他のリソースとの連携手法について，近年の Web 技術やデジタル・ヒューマニティーズの手法の流れの中に位置づけつつ報告し，テキスト・データベースの新たな可能性について検討する．

Exploring the Potential of Text Databases through Collaboration with Other Resources

Kiyonori Nagasaki† Toru Tomabechi† Masahiro Shimoda††

The SAT2012 is a new version of a Web service that delivers online Buddhist scriptures along with various cooperative resources as a way of bringing forth the optimum possibilities of a textual database. As the result of a collaborative effort, the SAT database is now seamlessly interacting with several online projects that deal with CJK ideographs, as well as field-oriented bibliographical databases. This paper describes the technical aspects and management of the collaborative works of the SAT2012 by comparing it with recent trends in current Web technologies and the digital humanities, at the same time discussing the possibilities of textual databases in general.

1. はじめに

SAT2012[a]は，SAT 大蔵経テキストデータベース研究会（代表：下田正弘．以下，SAT）が 2012 年 6 月に公開した，大正新脩大蔵經テキストデータベースの Web サービスの新しいバージョンである．本稿では，この SAT2012 の構築において取り組んだ様々な問題とその解決策について論じることで，テキスト・データベースの新たな可能性について検討したい．

SAT では，1994 年より，85 巻，約 1 億 5000 字にのぼる主に漢訳の仏典である大正新脩大藏經[1]のテキスト・データベースの構築に取り組んできた．大正新脩大藏經は大正末期～昭和初期にかけて活版で印刷された洋装本の仏典シリーズであり，それまでの仏教学の成果を集約した決定版として世界中に流布し，多くの大学図書館等に所蔵され，現在も基本文献として広く参照されているものである．それゆえ，テキスト・データベース化することの意義はきわめて大きく，大きな期待を受けて開始された取り組みであった．このプロジェクトは，対象テキストの分量の多さだけでなく，漢文であったことや当時の活版印刷にしばしば見られる異体字の問題などから，テキスト入力自体が困難を極めた．プロジェクト開始の頃には E-mail が一般化し始め，続いては Web が広まっていく時期であったことから，IT 環境の変化にあわせて作業環境も変化していき，2005 年には Web コラボレーション上で構築作業が行われるようになり[2]，2007 年 7 月にようやく完成に至り，2008 年 4 月には Web サービス版（以下，この版を SAT2008 と呼ぶ）をも公開することとなった．

2. SAT2008 について

SAT2008 についてはすでに様々な形で論じてきているため，詳細はそちらを参照していただきたい[3]．基本的な構造としては，全文検索を容易に行えるようにした上で，検索された本文を読む際に，それを支援するための参考資料を参照しやすいように工夫を凝らすというものであった．

2008 年の公開当初より，佛教用語の英訳を提供する電子佛教辞典(DDB)[b][4]の公開データを活用し，英語による可読性を高め，国際的な利用者への利便性を高める工夫を行っていた．具体的には，漢文のテキストをドラッグするとそれに対応する英訳が表示されるというものであった．

Fig. 1: DDB の検索結果画面

† 一般財団法人人文情報学研究所
†† 東京大学大学院人文社会系研究科
a) http://21dzk.l.u-tokyo.ac.jp/SAT/
b) http://buddhism-dict.net/ddb/index.html

また, DDB の見出し語との最長一致による対応付けにより英訳を表示することでより読みやすさを高めていた. (Fig. 1)

一方，本文全文検索に際しても，英語による利用者に配慮すべく，検索語入力欄に入力した英単語から DDB を通じて漢訳の検索語候補を列挙してクリックのみで選択・検索できる機能も用意されていた．また，日本印度学仏教学会が 1980 年代より継続的に構築運営しているインド学仏教学分野論文書誌データベース INBUDS[c]との連携機能も提供されていた．これは，CiNii WebAPI 公開後，それを利用して CiNii[d]における論文 PDF の有無を確認・表示し，論文 PDF があれば CiNii の当該頁にリンクする機能も追加され，二次資料へのアクセスをより容易にした．また，この間，筆者らは，国内外の様々な Digital Humanities 関連・仏教学関連の研究集会にて発表して意見交換を行い，それを反映すべく改良に取り組みつつ，SAT2012 となる大規模改修へ向けて検討を続けた．一方，研究用途を損なわない範囲での改良ということで，韓国の仏教者の間では，仏教用語の発音をハングルで書くことはできるが，漢字の入力の方法がよくわからず検索語の入力がやや困難であるという問題があることを知る機会があったため，これも，DDB のハングルの読みの情報を用い，上述の英訳語から仏教用語を引くのと同じ仕組みを用いて，ハングルの読みから仏教用語を引き，検索できる機能をも提供した．

3. SAT2012 について

3.1 開発に関して

SAT2008 では，上述のように，ユーザビリティを高めるために小さな改良を重ねてきていたが，後述するいくつかの大規模な機能を新たに追加するためには，抜本的にインターフェイスを作り直す必要が出てきていた．特に，SAT2008 では頁をすべて読み込み直すアクションが多かったが，これでは，大規模な機能追加をした場合には毎回大量のデータを読み込むことになってしまい，ユーザビリティの面で不都合が大きく，一方で，AJAX を駆使し頁遷移を減らすことで読み込みデータ量を減らし筒スタンドアロンのアプリケーションのようなより良いユーザビリティを提供することが可能となりつつあったため，そのような方向を目指した抜本的な改良版の作成を進めつつあった．

上述の小改良についても抜本的な改良についても，結果的には，IT 企業にプログラミングを外注することなく，また，既存の CMS 等のシステムを援用することもなく，内製で開発を行うことにした．内製での開発は，開発の継続性や開発者の作業負担，さらには開発者自身の知識・技術による制約ということが問題として挙げられることが多い．そうした問題を避けるため，外注や，既存の CMS 等のシステムを少しのカスタマイズで利用することはしばしば行われ，良い成果を挙げている場合も少なくない．とりわけ近年では，いわゆるメタデータ CMS である Omeka[e]を中心として，Neatline[f][5]や Scripto[g]といった，コラボレーションをも視野にいれた優れたフリーのシステムが登場してきており，また，CMS の代表格の一つであり Wikipedia で使われている MediaWiki も様々な便利なプラグインが提供され，ユーザの多様なニーズに対応できるようになってきている．しかし一方で，まず，外注に関しては，研究者・利用者のニーズに細かく対応した作り込みを含めた仕様を適切に策定することは容易ではなく，リリース後の Web サービス環境の変化に対応した小改良を依頼するにもその都度少なくない費用がかかり，さらに，大規模改修の際には相当な費用がかかること，そして，技術の継続性ということについても実際にはあまり期待できない場合もあることが経験的に予想されたことから，研究者兼開発者という人材が確保できる状況では，ニーズを具体的に知っている者が自分で開発を行う方が時間的にも費用的にも効率的であると判断し，SAT2012 に関しては引き続き永崎が開発を行った[h]. また, CMS 等の既存のシステムの利用に関しては，SAT2012 での要件を満たしつつ，これまでの資産を可能な限り継承するということを方針としていたため，手頃な CMS のようなものが見つからず, それまでと同様に Apache, PostgreSQL, PHP, jQuery を用い，サーバサイドは PHP，クライアントサイドは Javascript で開発を行った．

3.2 データ作成に関して

SAT2012 では，SAT2008 に対して新規のデータを追加した．その内容は，頁画像，英訳大蔵経と大正蔵とのパラレルコーパス，校正済みテキストであった．これらはいずれも，外注することなく，大学院生やオーバードクター等の関連分野の若手研究者の手によって行った．大きな方針としては，人文系分野ではデジタル化関連の仕事に限らずしばしば行われていることだが，若手研究者に少しでも研究と近いところでアルバイトをする環境を提供することで，研究分野全体の発展に少しでも貢献することを目指すというものである[i]．特に，パラレルコーパスについてはある程度のテキストの読み込みが必要となるため，研究者としての訓練や情報収集という意味でも相応の意義がある．また, パラレルコーパス構築及びテキストの校正に関しては，Web コラボレーションシステムを採用したワークフローを用意することで，若手研究者が自宅等からでも比較的無理なく参加できるよう配慮している．また，若手研究者がこ

c) http://www.inbuds.net/
d) http://ci.nii.ac.jp/

e) http://omeka.org/
f) http://neatline.org/
g) http://scripto.org/
h) 内製の比較的成功している例としては東京国立博物館の e 国宝や佛教大学図書館デジタルコレクション[6]などがある.
i) たとえば，立命館大学 GCOE 日本文化デジタル・ヒューマニティーズ拠点において提唱・展開されていた ARC モデルもこのような営みの一環として捉えることができるだろう

のようなデジタル化に関わる仕事に参加した際に，成果を自身の研究成果につなげるという試みも一部で行われている[7]. 研究成果にまでつながることはより望ましいことだが，相応の負荷も発生する可能性が高く，若手研究者自身の研究と仕事の内容がかなり近いものである必要があるだろう. このことについては今後の検討課題としたい. また，研究者でない人も含めて広く協力者を求める Web コラボレーションも一部で始まっているが[8]，仏教学の場合にも，仏教学を大学・大学院で修めた後に社会で活躍している人々が少なくないため，そういった人々の協力を得られるような枠組みを検討していきたいと考えている．

パラレルコーパスの構造や特長，その意義についてはすでに各所で報告してきているため[9]，ここでは簡潔に触れておく．パラレルコーパスは，漢文と英文の対応付けとして作成されるため，まず，仏教における漢訳語を英訳する際の用例，あるいは，英語から漢訳仏教用語の用例を確認することができる．世界中の研究者だけでなく，仏教者にとっても，英語で仏教に関する議論をする際にこの情報が有益となることは疑いない．それのみならず，古典文献では句読点の位置情報が必ずしも絶対的なものではないということは，1950 年代のテキストの統計処理研究においてすでに指摘されており[10]，大正新脩大藏經についても，句読点の位置，すなわち，文章の区切りがあくまでも校訂者の解釈を反映したものであるに過ぎず，さらに，句読点の間違いも散見されるため，文章の区切りに関する情報がデジタル翻刻の過程で十分に得られていない．一方，これに基づいて英訳された財団法人仏教伝道教会の英訳大蔵経では，それぞれに翻訳者の方針に基づいて文章が区切られている．英訳と漢文のパラレルコーパスを英文の文章を単位として作成することで，漢文に関しても一貫した方針に基づく文章の区切りが提示されることになるのであり，このことは，漢訳仏典を人が読む際に参考になるだけでなく，機械処理をする際の基盤としても有益なものとなることが期待される．また，この対応付けは，言語が異なるためにオーバーラップする場合があり，TEI エンコーディングに落とし込む際にはいわゆる Stand-off マークアップ[11]として扱うことになる．Stand-off マークアップに関しては，たとえ oXygen[j]のようなユーザビリティの高い XML Editor であっても，人が直接タグを扱うには複雑過ぎると見られており，CATMA[k]や JUXTA[l]等，ユーザにタグを意識させないようにするというアプローチが一つの潮流となりつつあるが，このコラボレーションシステムも，ユーザに XML タグをまったく意識させずに Stand-off マークアップを作成できるという点では，その潮流の一つとして位置づけることができる．対応付けの具体的な手法については，校合資料を扱う際に用いられている 3 つの対照手法のうち，Location-referenced method[12]を採用する形で試行している．これは直接 XML を扱う場合にはやや複雑で手間がかかるが，機械処理によってデータを作り出せる場合には複雑さに関しては問題ない．ただし，他の多くの TEI エンコーディングの手法と同様，より多くのアプリケーションの登場が期待される状況であり，データを増やしていくことで国際的にアプリケーション開発の機運を高めていく必要があるだろう．

このパラレルコーパスの構築に関しては，これまでに 19 人の若手研究者が参加し，約 30000 件の文章単位の英文-漢文パラレル情報が集積されており，現在も継続して構築され続けている．なお，構築過程では，参加者からのフィードバックによるシステムの改良が行われ，とりわけ，パラレル情報の確認と校正に関わる管理システムの効率改善に関わる改良に多くの労力が注がれた．このようなサイクルによって研究分野でデジタル化に関わる知見が効率的に蓄積されることは，システムからデータ作成までのすべてを研究分野内で完結させることの一つの大きな意義であると言えよう．

テキストの校正に関しては，SAT2008 では 1994 年のプロジェクト開始当初には Shift-JIS をベースとしつつ検索の便のために字形を統合するといったことを行ってきたものを，2006 年頃から字形になるべく忠実に作成した上で字形の違いは検索時点で包摂できるようにするという枠組みに変更し，さらに，2008 年に UTF-8 を利用する方針へと移行したことから，それらの方針が十分に徹底されておらず，さらに，いわゆる Shift-JIS の外字を UCS 符号化文字と対照して UCS 外字かどうかを判断する必要が生じていた，といったような事情により行われることになった．校正作業用の Web コラボレーションシステムでは，頁画像を Web 上で閲覧しつつ校正を行える仕組みを用意することで校正を効率的に行えるようにした．これに関しては，謝金を計算するための作業開始終了時点での Web 打刻を含めた管理システムについて多くの改良を行い，今後何らかの形で報告したいと考えている．校正の内容としては，結果として，これまでのデジタル翻刻時のミスも同時に行われることとなっている．また，UCS 外字として扱わざるを得ないと判定されたものが 6000 字程度となっており，これらのうち 3000 字程度は CJK 統合漢字拡張 F として UCS 符号化提案を行うということで，IRG に提案されたところである[m]. この符号化提案についての詳細は別に報告する．

3.3 システムに関して

SAT2012 のシステムに関して若干触れておこう．サーバコンピュータは東京大学文学部の LAN に設置されており，CPU は 4 コアで 2.13GHz の Intel Xeon E5606 を 2 基，DRAM

j) http://oxygenxml.com/
k) http://www.catma.de/
l) http://juxtacommons.org/

m) http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg39/IRG39.htm

は 32GB となっている．ハードディスクは内蔵 600GB×2（RAID1）のものに OS やデータベースを載せており，加えて，頁画像のデータを外付け HDD ストレージ（物理容量計 8TB, RAID5）に載せており，これが約 1.8TB を占めている．OS は Red Hat Enterprise Linux Server release 6.2 であり，これに上述の各種サーバソフトウェア類をインストールして稼働させている．なお，上述の各種コラボレーションシステムについては別のネットワークに設置されたサーバ上で運用しており，システム構成はほぼ同様である．

3.4 検索に関して

SAT2012 では，全文検索に際しては，全文検索ライブラリである Senna[n]による n-gram インデックスを用い，さらに，オプションとして，(1)互換漢字の正規化をしない(2)英数字は単語ごとではなく文字毎にインデックス化する，という機能を用いている．(1)については，自前で作成した検索システムの方で CHISE[o]のオントロジーを適用して漢字の曖昧検索機能を提供しつつ厳格な検索機能も提供しているため，全文検索ライブラリでは正規化をしない検索のみを実行できれば十分だからである．(2)に関しては，単語ごとでインデックス化してしまうと，外字番号（&MT[0-9]{5};という書式を用いている）を検索することができない場合があるためである．検索コストが若干増えてしまうが，使い勝手の面では体感的に速度が変わることはなかったため，このような形で設定している．なお，SAT2012 で 85 巻分すべてを全文検索した場合に要する時間の目安は現時点では以下の通りである．(Table. 1)

検索語	所要時間	ヒット件数
普賢菩薩	1 秒未満	2,438 件
阿修羅	1 秒未満	5,352 件
文殊	約 1 秒	23,577 件
菩薩	約 7 秒	346,318 件
仏（「佛」も自動曖昧検索）	約 18 秒	759,828 件

Table 1. 検索語別検索所要時間

テキスト検索関連の設定については特にチューニングは行っておらず, 若干の高速化の余地はあると考えられる．しかし，検索速度向上についての要望は現在のところユーザからは出ておらず，大正新脩大藏經で設定されている分類ごとに絞り込んでの検索や, 巻号毎に絞り込んでの検索, あるいは各経典単位での絞り込み検索も可能となっていること等から，検索の高速化に関しては当面は必要性がそれほど高くないということが考えられる．検索に関する要望としては，SAT2008 の頃に KWIC 検索をしたいという要望が複数のユーザからあったため，SAT2012 では検索に関する様々な設定を用意する中で KWIC 検索機能を選択できるようにした．このほか，本文のみの検索，脚注のみの検索が可能となっている．また，正規表現検索やワイルドカード検索のようなことをしたいというユーザがいるが，検索速度やセキュリティの問題をクリアする必要があるため，これについては対応を検討している段階である．

3.5 頁画像とのリンク

SAT2012 の大きな改良点として，大蔵出版株式会社の許諾の下，大正新脩大藏經の頁画像がすべて公開されたという点が挙げられる．ただし，大蔵出版は現在も大正新脩大藏經を販売しており，丸ごとすべての頁画像をダウンロードできる形にしてしまうことはややはばかられたため，画像を分割して表示し，頁の全体像を拡大して見ることはできないようにしてある．それだけでは読みづらいので，本文中に 10 行ごとにリンク画像を用意し, それをクリックすると当該箇所とその周辺が表示されるようにしている．これは ImageMagick[p]を用いて事前にすべての画像を分割し, 縮小版と拡大版を用意した上で, 独自に作成した Javascript のプログラムを使って読み出している．縮小と拡大は jQuery-UI[q]のスライダー機能を使ったインターフェイスを用意し，倍率に応じて縮小版か拡大版をサーバから読み出すようになっている. 画像は 600dpi でスキャンしており, 大正新脩大藏經の活字であればある程度までは詳細に確認可能である．

また，SAT のテキストデータは頁番号・段落番号・行番号がすべてのテキストに付されていることから，頁画像上の位置とある程度の対応付けが可能である．そこで，すでに SAT2008 において用意されていた,「テキスト番号・巻番号・頁番号・段落番号・行番号を URL で指定すると当該箇所頁が表示され当該箇所までスクロールする」という機能にこの頁画像表示機能を組み合わせ，特定箇所のテキストを表示すると同時に，頁画像も指定された該当箇所を中心として表示されるようにした．これにより，URL でテキストのある箇所を指定すると，電子テキストが表示されると同時にその頁画像の該当箇所も表示されるということになり，利便性をより高めることができた．

3.6 インターフェイス全体に関して

SAT2012 では，多岐にわたる連携機能を付加したため，一つの画面ですべてを表示することは困難となってしまった．そこで，画面と機能を大幅に整理した．画面を左・中央・右の三つに分け，さらに，左ウインドウはタブを用いて 3 つの画面を切り替えられるようにし，それぞれ，大正新脩大藏經の目次，ツールウインドウ（後述）制御，頁画像を配置した．中央ウインドウは div エレメントで検索結果表示，本文表示を配置しつつ，ユーザの動作にあわせて表示・非表示（show/hide）が切り替わるようにした．そし

n) http://qwik.jp/senna/FrontPageJ.html
o) http://chise.zinbun.kyoto-u.ac.jp/
p) http://www.imagemagick.org/script/index.php
q) http://jqueryui.com/

て，右ウインドウはツールウインドウを配置するための空白領域とした．すなわち，ユーザが同時に遣う必要があるものとないものを分け，それに応じての配置を行ったのである．また，特に，他のシステムとの連携機能等の付加的な機能については，各機能に jQuery-UI の dialog 機能を用いたフローティングウインドウを用意しておき，ツールウインドウ制御画面 (Fig. 2)にて，ユーザが必要に応じて適宜それらを表示・非表示にできるようにし，また，表示位置を自由に動かすこともできるようにしている．

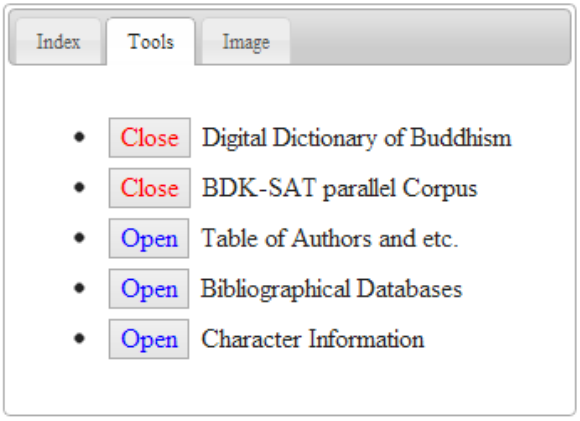

Fig. 2 フローティングウインドウの制御画面

以下に紹介する連携機能は，いずれもそれらのフローティングウインドウ上で動作するものである．

3.7 他のシステムとの連携に関して

DDB，及び INBUDS との連携に関しては SAT2008 のものをそのまま引き継いでおり，すでに述べたとおりである．

Fig. 3 パラレルコーパスのフローティングウインドウ

また，上述の英訳大蔵経とのパラレルコーパスに関しては，テキストの任意の一部をドラッグすると，その語が含まれるパラレルコーパスがフローティングウインドウに表示されるようになっており，これをもって用例の確認ができる．（Fig. 3）さらに，このウインドウ内の漢文テキストをクリックすると当該経典が表示され，当該箇所までスクロールするようになっている．また，英訳と直接紐付けされているテキストの場合には，ドラッグした箇所に紐付けされた英訳が別のウインドウで表示されるようになっている．これにより，漢文仏典の英訳の仕方について様々な情報をユーザが効率的に得られるようになっており，この機能は世界中のユーザから好評を得ている．

書誌情報データベースとの連携に関しては，INBUDS との連携以外に，新たに CiNii と SARDS[r]との連携を組み込んだ．いずれも，テキストをドラッグすると検索窓にキーワードが入り，検索ボタンをクリックすると別ウインドウが開いて検索結果が表示されるというものだが，SARDS はドイツのハレ大学が中心となって構築されたインド学の欧文研究書・研究論文の書誌データベースであり 6 万件を超える情報が集積されている．ただし，漢字の情報はローマ字表記となっているため，漢訳語そのままでは検索してもあまり意味がないことから，SARDS との連携機能に関しては，DDB のリストを経由させることで，英訳語やサンスクリットのキーワードが検索窓に入力されるようになっている．CiNii に関しては，INBUDS での論文情報検索では分野横断的な検索ができず CiNii を使えた方が便利な場合があるというユーザの要望に応えて連携検索機能を導入した．

文字関連の情報を参照しやすいようにする機能も提供している．これは校正作業において特に必要となったことから校正用システム向けに開発したものだが，さらに改良を加えたものを SAT2012 にも転用した．(Fig. 4)

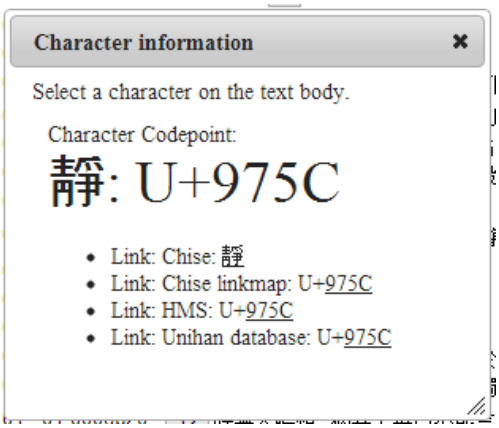

Fig.4 文字情報のフローティングウインドウ

機能としては，本文上でドラッグした 1 文字の UCS での

r) http://www.indologie.uni-halle.de/sards/

コードポイントを表示するとともに，そこから，CHISE，CHISE Linkmap，Han Morphism System，Unihan，における当該文字の情報へのリンクが生成され，クリックすると別ウインドウで表示されるようになっている．なお，コードポイントの取得は Javascript で行っているが，留意すべき点として，Javascript は内部コードが UTF-16 となっており，サロゲートペアに対応するためのスクリプトを自前で用意する必要があったことには留意しておきたい．

3.8 履歴機能に関して

SAT2008 への改善の要望として，前の画面にうまく戻れるようにしてもらいたいというものがあった．このような要望が出てくる背景には，一部のアクションに AJAX を利用していたため前の画面に戻りにくい場合があるというシステム上の問題と，新規タブや新規ウインドウを立ち上げることを好まないというユーザの習慣とが混在していたようであった．しかし，SAT2012 の場合には，より効率的に様々な機能を利用できるようにするために AJAX を多用して頁遷移を減らしたことで，ブラウザの「戻る」ボタンを使って以前の画面に戻るという一般的なアクションがほとんどできなくなってしまった．そこで，利用者の便を図るため，履歴機能を提供することとした．具体的には，クッキーを利用して，頁を見る，検索をする，検索結果をたどる，といった基本的な動作に関して記録を残しておき，それらをフローティングウインドウにリスト表示してリンクとしてたどって戻ることができるようになっている．(Fig. 5)

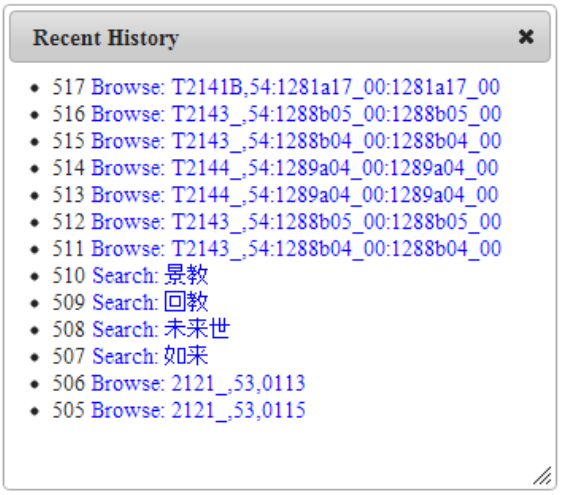

Fig. 5 履歴機能のウインドウ

4. ユーザとのつながり : 新たな可能性に向けて

SAT2012 に限らず，いわゆる人文系データベースは対象ユーザがある程度明確なものであり，ユーザにとってどのように有益であるかということは一つの重要な指標となる．また，一方で，このように有益であるべき，といった基準のようなものについては，一般的な Web ユーザビリティに関する基準は存在しており [s]，デジタル化されたテキストという観点からは TEI 等においてそうした指標が用意されつつあり様々な動きがあるが，人文系データベースとしての Web サイトに必要とされる固有のユーザビリティ，もしくはその構築の手法に関しては，今後しばらくの試行錯誤と議論が必要であると思われる．最後に，このことについての検討を行い，本稿の結びとしたい．

ユーザが SAT2012 は Web で公開しているものであり，アクセスログ等からある程度ユーザの動向が想定可能である．SAT2012 公開後，7 月からの 5 ヶ月間の延べアクセス件数は，以下のようになっている（Fig. 6）．

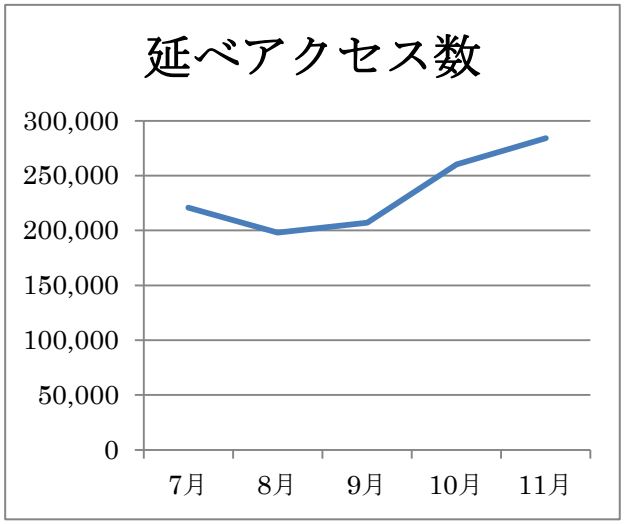

Fig. 6 2012 年 7～11 月の SAT2012 への延べアクセス数

これを見る限りでは，ユーザが順調に増えつつ利用回数も増えているということが想定されるが，一方で，卒業論文・修士論文等での利用にあわせてのびているということも考えられる．引き続き，利用傾向には注目していきたい．また，クッキー情報から得られるユニークブラウザ数は 23,052 となっており，日本印度学仏教学会の会員が約 2400 名であることを考えると，海外の研究者や他分野の研究者，一般ユーザ等にも広く使われていることが予想されるとともに，ユーザが様々な機能を利用していることが想定される．なお，海外からの利用に関しては，この 5 ヶ月間のアクセスログからトップレベルドメインを見た限りでは，上位 5 ドメインで 105,4593 件のアクセスとなっており，edu ドメインはほぼ確実に米国からのアクセスと言えるが，その件数は 33,958 件となっている．com ドメインや net ドメインの一部も米国であり得ることを考えるとかなりのアクセスが米国から行われているとみられる（Fig. 7）．また，6 位以下のアクセスについては，20 位までをあわせて 44,044 件となっており，その内訳は（Fig. 8）に示すとおりである．

s) http://www.usability.gov/guidelines/index.html

仏教研究の国際的な広がりが利用状況に反映されていると言えるが，SAT では，SAT2008 公開の頃より国際的な利用の広がりを考慮して基本的に英語によるインターフェイスの提供をしてきており，SAT2012 でもそれを踏襲し，インターフェイスだけでなく詳細なマニュアルをも英文で提供している．その配慮の有効性がこのアクセスログに現れていると言うこともできるだろう．

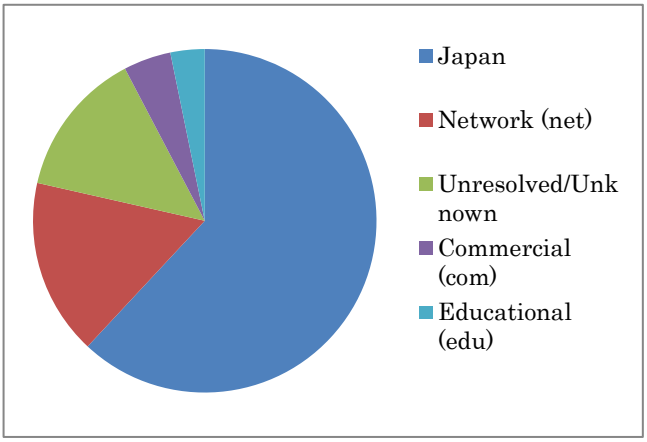

Fig. 7　2012 年 7 月～11 月の TLD 別アクセス数第 1～5 位

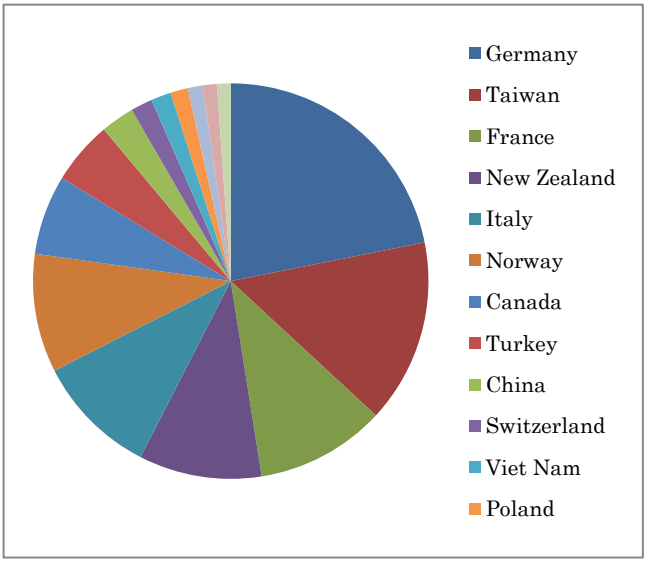

Fig. 8　2012 年 7 月～11 月の TLD 別アクセス数第 6～20 位

これまで，SAT2008 では，アクセスログの解析から得られる情報とともに，可能な限り様々な場でユーザの声を集め，それらを分析した上で改良を重ね，それを SAT2012 の開発の際にも反映させてきた．しかしながら，この方法だけでは限界があり，より広くユーザのニーズを探し出して反映させる必要が感じられたことから，SAT2012 の多様な機能を適切に周知することも兼ねて，各地でユーザ講習会を開催する方針を決めた．すでに第一回を 2012 年 11 月 19 日に北海道大学文学部にて開催し，仏教学のみならず国語学等の他分野の研究者も含めて 26 名の参加者が参集した．詳細な利用方法の周知が行われ，質疑応答では，実際の利用方法に基づく具体的な機能追加の要望から短期的・長期的なデジタル化資料のあり方についての議論に至るまで，人文学におけるデジタル化への期待の高さを如実にあらわすものとなった．Digital Humanities をはじめとする人文学におけるデジタル技術応用の営みにおいては，様々な形でデータベースを作ることやそれを用いて研究成果を出すような活動は盛んに行われているが，作られたデータベースがどのように評価され，そこからどのような改良が行われるかという枠組みについてはまだあまり議論が行われておらず手法も確立しているとは言えない．国立国語研究所における現代日本語書き言葉均衡コーパス(BCCWJ)を中心とした共同研究班による一連のワークショップ等[t]はその意味でも注目されるところだが，SAT においてユーザとの新たな対話の場を形成しようとする試みもまた，人文系データベースのより適切な構築と運用の枠組みを創り出すことに貢献できればと考えている．

謝辞　SAT2012 の公開にあたっては，多くの研究者の方々のお力添えをいただいたことを記して感謝したい．とりわけ，実際の作業に携わってくださり様々な改良案をくださった若手研究者の方々の助力なしには SAT2012 がこのような形で公開されることはなかっただろう．また，（財）全日本仏教会及び（財）仏教伝道教会にはそれぞれに，校正作業，及び BDK-SAT パラレルコーパスの構築に関して貴重なご支援をいただいた．なお，本稿で報告されている研究成果には，JSPS 科研費 22242002 の助成を受けたもの，JSPS 科研費 60601680 の助成を受けたもの及び JSPS 科研費 30343429 の助成を受けたものが含まれている．

参考文献

1) 高楠順次郎編『大正新脩大藏經』大正新脩大藏經刊行会, 1924-1934.
2) 永崎研宣, 鈴木隆泰, 下田正弘「大正新脩大藏經テキストデータベース構築のためのコラボレーションシステムの開発」『情報処理学会研究報告』 CH-70 (2006 年 5 月), pp. 33-40.
3) Kiyonori Nagasaki, A. Charles Muller, Masahiro Shimoda: Aspects of the Interoperability in the Digital Humanities, *Digital Humanities 2009*, 2009 , pp. 375-377.
4) A. Charles Muller, Kiyonori Nagasaki and Jean Soulat: The XML-Based DDB: The DDB Document Structure and the P5 Dictionary Module; New Developments of DDB Interoperation and Access, *Chung-Hwa Buddhist Journal*, vol. 25, pp. 105-128.
5) Bethany Nowviskie, et al.: Geo-Temporal Interpretation of Archival Collections Using Neatline, *Digital Humanities 2012*, 2012, pp. 299-302.
6) 飯野勝則「佛教大学図書館デジタルコレクションの設計とデザイン」『カレントアウェアネス-E』No.219, 2012, http://current.ndl.go.jp/e1315 (2012 年 12 月 21 日閲覧).
7) Constance Crompton and Raymond Siemens: The Social Edition: Scholarly Editing Across Communities, *Digital Humanities 2012*, 2012, pp. 441-443.

t) http://www.ninjal.ac.jp/research/project/a/sousei/sousei-pm/

8) Melissa Terras: Present, not voting: Digital Humanities in the Panopticon: closing plenary speech, Digital Humanities 2010, *Literary and Linguistic Computing*, Vol. 26, Issue 3, pp. 257-269.
9) Kiyonori Nagasaki, Toru Tomabechi and Masahiro Shimoda: Toward a Digital Research Environment for Buddhist Studies, *Digital Humanities 2011*, 2011 , pp. 342-343.
10) 三上 悠紀夫「古典校訂本の句点は信用できるか」『計量国語学』Vol. 9, 1959, pp. 22-24.
11) Lou Burnard and Syd Bauman ed**.:** Non-hierarchical Structures**,** *P5: Guidelines for Electronic Text Encoding and Interchange***,** http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html (2012 年 12 月 21 日閲覧).
12) Lou Burnard and Syd Bauman ed.: Critical Apparatus**,** *ibid*. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html (2012 年 12 月 21 日閲覧).