

# 「歴史学用ツール SMART-GS : 文書画像のサーチ・マークアップ・リンク」

林 晋 (京都大学大学院文学研究科 教授)

## 概要

文献研究用のツール SMART-GS の紹介をする。ハードコピー文献の分析を行うとき、我々は重要部分に下線を引いたり、丸で囲んだり、余白に書き込みをしたりなどして、文献の構造や内容を明らかにし、また自分の解釈を構築していく。これはHTML文書などのマークアップに対応する作業と考えることができる。HTML文書やPDF文書のような電子文書では、同様な作業が可能だけでなく、さらにサーチが可能であり、またマークアップ間にリンクをはることも可能である。

このようなマークアップ、リンク、サーチが、手書きなどの歴史研究用の文書のデジタル画像に対して可能になれば、それが歴史研究におよぼす利便性は大きい。実は100ページ程度の手書き文献に対して、SMART-GS はすでにその機能を実現しており、講演者の研究室では実際の歴史研究に日常的に使用されている。

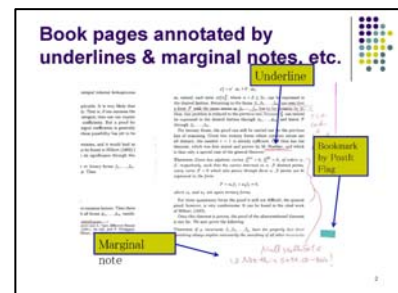
同じことがWWW上でもできる状況を想像して欲しい。たとえば、WWW上の電子アーカイブの史料に自由にマークアップ、リンク、サーチができたらどんなに便利だろうか。また、地球の裏側の歴史家と、同じ一次資料を共有して、日常的に共同研究ができたらどれだけ素晴らしいだろうか。実は、そのような環境を開発するために必要な基本情報技術はすでに存在するのである。

## はじめに

ご紹介にあずかりました林です。三上先生のほうから私を歴史家としてご紹介いただきまして、大変喜んでおります。もともとは20年ほど工学系の学部におりました、コンピューターサイエンスとか、ソフトウェアが専門の者です。歴史は子供のころからの大好きなサブジェクトでいろいろやっていたら、こういうことをやるようになりました。実はここでお見せするツールも、もともとはソフトウェア開発のためにつくったものがベースになっていますが、今日お話しいたしますのは、SMART-GS という歴史用ツールで、今ありましたお話とも非常に関連が深いのではないかと思います。これは私の京大文学研究科の私の研究室と、北大の情報科学研究科のMeme Media Laboratory との共同研究です。

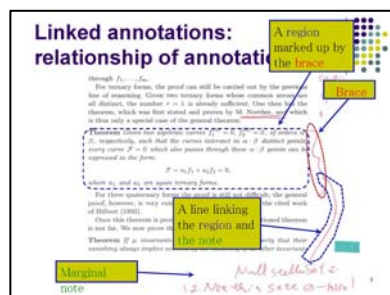
## 文書史料を分析する1：マークアップ、コメント

このスライドは19世紀のドイツの数学者の1897年のレクチャーノートを英訳したものです。我々がこのような文書を読むときにどうするかといいますと、例えばここにM. Noether という別の数学者の名前が出ておりますが、ここにアンダーラインが引いてある。これは一種のマークアップとすることができます。このページ全体にポスト・イットフラッグという3Mがついているものをつけまして、これはページについてのマークアップ。さらにはマージナルノート (Marginal Note)、この空白のようなテキストで自分の意見を書くわけです。先ほどですとRDFみたいなものです。こうやって我々はマークアップをしながら本を読んでいく、あるいは歴史史料を分析をしていくわけです。



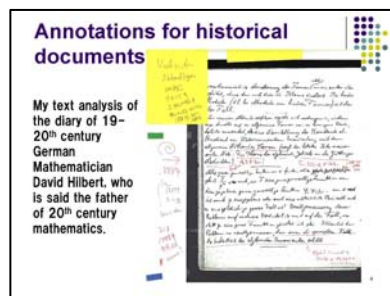
## 文書史料を分析する 2 : リンク

これは今の文書のクローズアップですが、ここに先ほどの下線があります。こういうふうなマークアップもいたします。右の Brace、日本語でいうと右中括弧 (}) を書いている。これは何かというと、要するにこういう領域 (Region) をマークアップしているわけです。そして、先ほどのマージナルノートがありまして、その二つの間にこういう線があります。これは何か。これはこの領域 (Region) とここのマージナルノートが関係があるということを示しているわけで、これは一種のリンクであるわけです。具体的に言いますと、実はこちらが主語、こちらが述語、先ほどの RDF でしたらこちらを主部にするような感じです。要するにこのラインがリンクとして、あるマークアップに対するステートメントを表現している。このようになります。



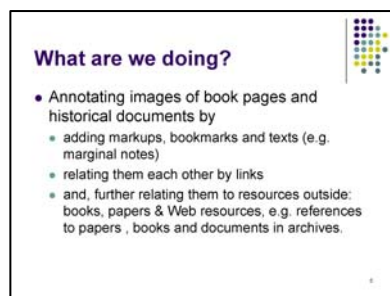
## 文書史料を分析する 3 : 手書き文献の場合も同じ

次のスライドは今のような印刷物ではなくても、1885年から1892年11月ころまで書かれた、数学者ダフィット・ヒルベルトのダイアリーの中から持ってきたものです。これはニーダーザクセン州・ゲッチンゲン大学図書館のマイクロフィルムを購入したものです。これを分析するとき、同じようにします。私は同じようにマージナルノートをここに付けています。それからマークアップもしております。先ほどと同じことですね。



## 文書史料の分析する : まとめ

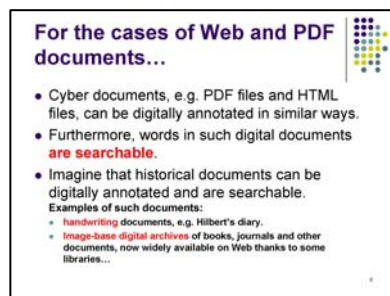
これらはどういうことをしているのか。まとめてみますと、こうなります。我々は本のページとか歴史史料をいろいろな種類のマークアップ、下線、リージョンを囲む、ページにブックマークをつける、さらにテキストを追加する。マージナルノートですね。そういうものを追加し、かつそれらをリンクによってコネクションをつけることによって、自分の考え、あるいは自分の Knowledge を表現している。さらにそのドキュメントの中だけではいろいろなものが足りない場合には、外側のリソース、本とか論文とかWEBリソース、こういうものに関する本の題名とページ数、WEBリソースでしたらURIでよろしいですね。こういうものをつけることによって、自分の思考を表現していくわけです。



## デジタル・テキストの場合にはサーチも可能

こういうことはデジタル文書の場合、PDFとかHTMLのWEB上のファイルに関しては当たり前のことになっております。PDFの文書に今のような Annotation とかリンクをつけられるというのはご存じだと思います。PDFの1カ所をカリカリッとやりますとブラウザーが立ち上がって、適当なところに飛ばしてくれる。それはリンクですが、これはもう当たり前のことです。

ところがPDFとかHTMLの場合、もう一つ要素があ



ります。サーチができる。これが便利なわけです。特に大量の文書になりますと大変便利です。もしこの1セット、つまり「マークアップ、リンク、テキストで自分のステートメントを入れる、サーチ」という機能のセットがあると大変便利であるわけです。もし、その1セットが、先ほどのような手書きの歴史文書、さらには Image Base Digital Archive というものが今ふえていますが、そういうものに対してできたら大変便利だとおもいます。その Image Base Digital Archive が何かといいますと、例で示します。国内ですと恐らくこれが最大です。

### 電子ライブラリーに応用できたら？

これがその例、国会国立図書館の近代デジタルライブラリーです。この前、大正時代の本が入りまして、明治から大正期の本が 14 万 3000 冊入っています。こういうふうにもとのデータを画像として撮りまして、目次だけキャラクターのテキストにしてあって、該当箇所にリンクが張ってあるというものです。いま私は明治時期のことを学生と一緒に調べておりますが、その研究に、これは本当に便利なものです。普通の図書館にはないものまであります。たとえば、坪内逍遙が行った論理学の講義などというものがあります。ただ、これは本体がないのですが、一応項目はあります。



こういうものに関しても、サーチができて、マークアップができて、リンクができて、テキストを追加できるとなったらどれだけ便利か？例えばこの近代電子ライブラリーに、自分のところではなくて電子ライブラリー上にマークアップをつけられるようになると考えてください。歴史に関する自分の考えを伝えるのに大変大きな力となるでしょう。しかも、別な学者が別な結論を主張して、それも書いて、かつお互いに批判などをトラックバックのように書けるようになったら、どうなるか。これは非常に楽しいのではないかと思います。

### SMART-GSプロジェクト

そういうものができたらいいな、特にネットワーク上でできたらいいなと考えて、これを見せると、1次史料と格闘している人は大きな興味をしめします。たとえば、ネットワーク対応にするというのは、私の同僚で仏教学の研究者からでた案です。仏教学はヨーロッパでも盛んですから、彼の研究者仲間はドイツにいるようです。



これはちょっと夢のような話に見えますが、実はネットワーク版はまだですが、ほかの部分は既に一応動いております。Java のアプリケーションで、これはその後ろに座っております私の元学生、いま東京で働いておりますが、小林和晶君が修士論文で作ったものが、今年の3月に一番最初のバージョンが動きまして、いま私と私の学生がヒルベルト研究に使うとともに、どんどん改善をしております。またネットワーク上でどう動かすかは、これからですが、仕組みだけはすでに考えてあります。これからネットワーク上に持っていかうとしてるところです。論より証拠といいますので、デモをごらんに入れます。

## 動画によるデモ 1 : サーチ

<http://www.shayashi.jp/SMART-GS/searchdemoForKahakuWS.wmv>

これはデモの様子を記録した動画で、やっていることはサーチです。この部分、これはヒルベルトの日記です。その中の、この部分が、レオポルト・クロネッカーという数学者の名前です。クロネッカーはヒルベルトの論敵として有名なので、この人に関してヒルベルトが何を言っているか、彼のダイアリーの中を全部見たい。ところがこの文書はドイツ人に見せても、よくわからん、読めないと言います。それぐらいの走り書きです。当然、日本人には読みづらい。探すのは大変です。そういう時に、これが使えます。

ここにあるクロネッカーの名前が赤く表示されています。これと同じようなつづりを探し出しますが、その検索プログラムが一番基本部分は、北海道大学の田中譲先生というところの技術です。赤石さんが出られたところですね。その研究室の寺沢さんが作ったものです。

さて、Search for the Images というところを押します。そうするとダイアログがでてきて、ここでどの範囲を探るかというのを、いろいろと指示できます。例えば画面の一つをコマと呼んでいます、ゼロ番のコマから 18 番のコマの画像を探せというのもできます。この実験ではそうではなくて、文書の全部から探せというので all spreads と書いてあるチェック・ボックスをチェックしています。で、サーチ。この動画の説明では 15 秒と書いてありますが、私のマシンでやりますと 15 秒よりもっとかかります。大きなマシンだと 15 秒でできるんでしょう。

今ここでごらんになっている、これがもとの「クロネッカー」です。そして、それとよく似たとつづりが沢山探されているわけです。ただし、このようにいっぱい間違えた答えも出ます。ここが普通のサーチと違うところです。これで精度が悪いから使い物にならないと思われていたんですが、「芋づるサーチ」というものを考えましたら十分実用になりました。これは、見つかった内の正しいものを人間が選んでどんどんためていって、ためているものの、どれかに似ているものを探し出す。そのプロセスを繰り返して探すと言う単純なやり方です。

たとえば、ここでいま見つかったもののリストを見ます。よく似ているように見えても、あまり筆跡が汚いので別なものが同じに見えることがあるので、文脈を見まして、これがクロネッカーかどうかを確認をし、そして正しいものには、「YES」というチェック・ボックスをクリックします。そうすると、私が「バケット」と呼んでいるものの中に、それが入ります。そうすると、次にイメージ・サーチをやると、今後はバケットに入っている二つのどちらかに近いものを探してくれるんです。二つでなく三つ以上バケットに入っても同様です。このようにして広げていきます。たったこれだけのアイデアで、精度がもう信じられないぐらい上がったんです。基本の技術は北大のもので、これだけが私の工夫があったところだと思います。

いま、2 回目のサーチの結果が表示されました。ここに新しく見つかったものが見えています。ちょっとここでごらんいただきたいのですが、ここに赤いものがあるのとないのがありますね。ないのは既に一度目のサーチで見つけたものです。この赤いのが新しいものです。そうすると当然新しいものを見たい。そういうときには、このように表示のオプションを切り替えます。

このような答えのリストは Google のように実は候補が、ずらーっと後ろに 100 も 200 個も並んでいるのですが、今は、そのトップの 40 ぐらいしか表示していません。(表示、オプションを切り替えて) これでいま新しく見つかったもののみを表示するようにしました。

そうすると、ちょっとこれはこの小さい画像ではよくわかりませんが、どうもこれとこれは当たっています。新しいのが見つかったわけです。これを再びバケットに入れて、これを2～3回繰り返します。大体全体にかかるのが80ページで5分から8分、ヒルベルトの日記の場合で、そんな感じです。

で、そのように繰り返しますとどのような結果になるかという、こういうふうになります。これは学生がやったものの記録ですが、これで8個ぐらい見つけています。私がやりますともう少し見つかります。このデモ動画を作った学生はドイツ語を知りませんので、何か見落としているものがあるんだろうと思います。

つまり、これは人間のスキルがコンピューターによって増幅されているのです。普通のサーチというのは完全にコンピューターにお任せですが、これは人間とマシンがMan-Machine Symbiosisを起こして、人間の歴史家としての、あるいは文献学者としてのスキルが増幅されているという、そういう意味でのサーチです。そういう意味では、今までのサーチとは全く違います。

### 動画によるデモ2：マークアップとリンク

<http://www.shayashi.jp/SMART-GS/MarkingUpAndLinking.avi>

さらにこういうサーチをやりましたとき、いろいろな新しいことがわかりますので、今度はそれを記録したいと思います。そのようなときには、このようなテキストの部分にマークアップをつけます。また、画像のほうにも当然マークアップをつけます。このテキストから画像にリンクを張りたいというので、コメントを入れます。何でここにコメントを入れるかは、後でご説明します。入れているコメントは、「これはテキストから画像へのリンクだ」という意味です。

同じように今度はこちらの画像から逆方向に、テキストに向かってリンクをはっています。ここで今度は逆に画像からテキストにリンクを張ることができます。

いろいろこういうふうにしてリンクをつけます。いろいろつきますと、自分がどんな風にリンクをはったのか全体を見たくなることがあるわけです。そのときにはReasoning Webと私が呼んでいるものを開きます。そうすると、こういうふうにならなくなったリンクが全部グラフになって出てきます。一番最初のテキストのマークアップ、そこから引いたリンク、テキストからイメージのリンク、そしてその先のイメージです。このイメージから引かれているリンクはイメージからテキストへのリンクであって、その先にはこのテキストがある。このようにして、歴史史料の様々な部分と自分が書いたテキストが完全に一つのWEBの中にオーガナイズされている様子の鳥瞰図が見えるわけです。

### 動画によるデモ3：SMART-GSのリンクはone-to-many

<http://www.shayashi.jp/SMART-GS/OneToManyLink.avi>

このシステムのリンクは、普通のWWWのリンクと違い、One-to-many Link になっています。一つの場所から複数の場所へリンクすることができます。例えば、このテキストのOne-to-many がマークアップされていて、そこから、こちらの画像の上のこのところにファーストリンクをはります。そのときコメントとしてファーストリンクと書きます。そして、別な適当なところをとりまして（黄色いマーカーでマークアップをつける）、マークアップには一応フェルトペンのようなマーカーがあります。こうやって黄色のマーカーで引くような形で、こちらの画像のこの部分にもマークアップをつけます。これですね。そして、ここへのリンクがセカンドリンク、そういうコメントを書いたリンクをはります。

こうやってやりますと、もとのOne-to-many のところでローカルビューというものを開きますと自分の行き先が2カ所見えますので、行きたいほうを見ます。右側に表示されて



いるのが、リンクにつけたコメントです。このリンク先は何だったかというのをリンクに付いたコメントに記録してありますから、たくさんリンクがあっても、どこに行くかはそれで判断できるということです。もっと複雑になって、わからなくなったら先ほどの Reasoning Web に帰って、あそこにもコメントを入れたり、見たりできるようにつくってありますので、それを利用してコメントを適当に参照したり入れたりしていけば博物館の中をブラウズするようにブラウズができるわけです。

#### 動画によるデモ 4 : 歴史研究に使った実例 1 林の数学基礎論史

<http://www.shayashi.jp/SMART-GS/FromHayashisResearch.avi>

先ほど同じような図が出ましたが、これは私が実際にヒルベルト研究のときにつくった Reasoning Web です。ただし、さっきのグラフだけの Reasoning Web と違い、そこにバックグラウンドとしてページ数のラインと年のラインを入れてあります。これは実はただ絵を持ってきて後ろに載せただけなんです。ただし、そうすると、このリンクで全部いろいろなことがわかってきます。このダイアリーには日付がありません。それをいろいろな Philological なテクニックを使いまして、例えば一番最後のをごらんに入れますと、あるノートには、ある論文の内容が未解決のように書いてある。そうすると、その論文のサブミッションより前にこのノートは書かれたとか、そういうことがわかるわけで、サブミッションの日付は記録が残っていますので、それでいつより前に書かれたか、荒い同定ができます。そういうふうにしてやりますと、大体このノートがいつごろ書かれたかわかるわけです。

こういう手法によって私は 19 世紀の数学基礎論史をいま実は書きかえつつあります。もしそちらの歴史のほうにご興味がありましたら、岩波文庫の『ゲーデル 不完全性定理』をごらんください。今までの数学基礎論の歴史を、かなりひっくり返しつつあると私は自負しております。

#### 動画によるデモ 5 : 歴史研究に使った実例 2 小林和博の卒論研究

[http://www.shayashi.jp/SMART-GS/SMART\\_GS\\_demo\\_kkobayashi.wmv](http://www.shayashi.jp/SMART-GS/SMART_GS_demo_kkobayashi.wmv)

もう 1 個ありますけれども、ちょっと時間がありませんので一瞬だけお見せします。これは今、私のプロジェクトで一番頑張ってくれている 4 回生の小林和博君が、コンピューターの HCI の技術史をやったもので、彼の卒論用です。卒論のための自分の知識を、こうやって Reasoning Web に載せているというものです。うちの研究室ではこういうふうの実用的に使われております。

#### 林の歴史研究の日常から

SMART-GS を使って、歴史研究をやるときのイメージは、こんな感じです。私の書斎にはワイドモニターが 3 台ありまして、マルチプルモニターにしてあります。これはその画像を一度に全部撮ったものです。真ん中だけちょっとモニターがでかいので、下が出っ張っております。これが SMART-GS です。ここら辺まで SMART-GS で、こちらがドイツのゲッチンゲン大学のデジタルアーカイブを広げているブラウザです。こちらは、Wikipedia で、ヒルベルトの顔です。こちらは 20 世紀の独英 (German-English) 辞典をデジタル化したものをオーストラリアの翻訳家が売ってしまっていて、それを買ったものです。ドイツ語というのはつづりがすごく安定していないのです。ですから、古い時代のものを使わないと古い文書は読めま



せん。これだけのものを使って私は研究をしているわけですが、これは非常に便利です。  
**ネットワーク版へ！**

このような便利な環境が、もしインターネット、WEB上でできたら楽しいですね。実は考えてみますと、これは割と簡単にできるのです。これはたまたま私がWWWに適応したように、SMART-GSの設計をやっていたんですね。私が自分で設計をしていたときには、ネットワークと言うことは何も考えていませんでした。ですから、これは偶然の賜物です。

### GS-file: アノテーションの方式

どういうことかという、実はこういうふうに画像データの上に赤い長方形のマークアップがあるということ、この仕組みはどうなっているかといいますと、ものすごく単純で、図の右のようになっています。この赤い長方形の座標をとって、歴史データは変わらないということを利用して、それに重ね合わせるだけです。

これがHTMLだとだめなんですね。ブラウザの横幅を変えますとテキストの改行の位置が崩れます。歴史文書の場合はこれが起きません。だから、こんな簡単なことができたんです。

今、ニコニコ動画というのと、それに類するWEBのサービスがありますが、WEBのサービスのほうが難しいんです。表示の形が見る人のブラウザにより変わるからです。

しかし動画は中々変えられないから、実はコメントを入れやすいんです。歴史文書は変わらないから、変えてはまずいですね。後ろに手が回ります。例の縄文土器を発掘したようなことになりますから。そうすると、その変わらないことを利用して逆に簡単にマークアップをするということが出来ます。

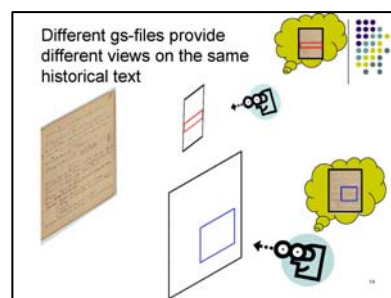
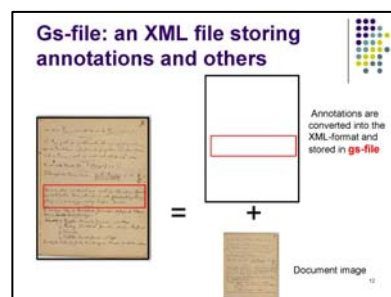
### 一つの史料に複数の解釈を同時に与える

しかも、これには色々良い「副作用」があります。どこかのサーバーにあるアーカイブに対して、ある学者と別な学者がそれぞれのマークアップをしていって、それを別な人が見るという状況を考えましょう。まあ、同じ人でもいいですが。そうすると、同じ歴史史料に対して、こういう風に幾つかのビューが同時にできて、しかももとのやつはみんなて共有している。こういう状況が簡単に作れます。そうしますと、例えば次のスライドのようなネットワークバージョンのモデルが考えられます。

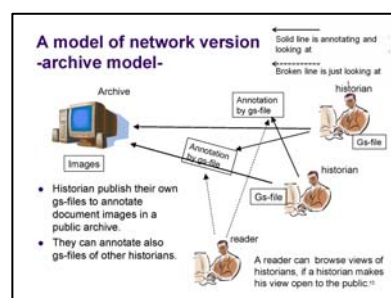
### ネットワーク版1：アーカイブモデル

このスライドの左上、ここにあるコンピューターの絵が、これが例えば国会図書館のアーカイブだと思ってください。ここのアーカイブの文書自体は、ユーザは何も変えませんが、変えることができないわけです。

それに対してある歴史家が、「あの福澤諭吉の文書の内容は、こうじゃ、こうじゃ」と意



見を公開したとします。ところが別な歴史家が、「いや、そうじゃない、私は、こうじゃ、こうじゃ、と思う」と反論と自分の意見を公開する。誰でもが、それを元の史料と重ね合わせてみれば見ることができて、さらに、お互いに相手の意見にも Annotation をつけて議論をすることができる。かつその議論をしているところを、また一般の読者が見ることができる。こういうものが割と簡単につくれるはずです。



## アーカイブモデルのモックアップ・イメージ

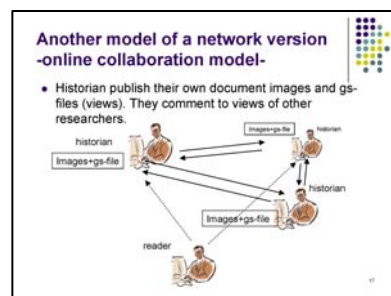
アーカイブモデルが現実になったらとしたら、どんな風にみえるか。例えばこれがモックアップ Mock-up です。つまり、Fake ですね。

ゲッチングンのアーカイブにこうやって勝手に自分でマークアップをつけて、みんなに見せる。こういうことが容易にできるはずです。



## ネットワーク版 2 : Peer to peerモデル

さらに考えますと、歴史家はどこかで自分で新しい歴史史料を見つけてくることがありますね。そのときにはアーカイブするのは大変ですから、自分のサーバーを持っていたら、そこでパブリッシュすることもできます。そういうものに対応するモデルを考えることもできます。このスライドがそのモデルです。つまり、SMART-GS の使い方について、いろいろな形を考えることができるわけです。



それならば、Peer to Peer の形、あるいはサーバー形式、あるいはローカルな PC 上、こういうものを全部統合したアーキテクチャーを考えよう。そのすべてを可能にしたプラットフォームを考えようというのが、これが SMART-GS プロジェクトです。ちなみに GS というのはドイツ語で Geschichte Studie、英語ですと History Study です。実は本当は杉田玄白のイニシャルなんですけど (笑)。

## 結論

結論です。SMART-GS という歴史用ツールがあり、ネットワーク版はまだありませんが、ローカル版はすでに存在し、私は実際に使っております。これはもう本当に重宝しております。たとえば、講義の前の日の夜中にヒルベルトの日記の何十ページもの手書き史料の中に Geometrie という単語がどのようにでてくるかを見たいというときがありました。

これは大変面倒な作業で、普通は数時間かかってもできるかどうかというタスクです。どうしようと思ったときに、SMART-GS でやってみたら 5分でできて、本当に自分でつくったツールに、自分で感激したことがあります。それからこれをもっと大きな、いま評判になっております元枢密院議長の倉富勇三郎の巨大な日記に適用しようという計画も、私の同僚の井永和先生が持ってい

**Conclusion**

- SMART-GS system has been already applied to a study of a very important diary in the history of mathematics and proved its usefulness.
- There is also a plan to use SMART-GS in a project to transcribe a **gigantic** diary of Yuuzaburo Kuratomi (1853-1948), a chairman of the privy council.
- The network versions in the future will be useful for research collaborations and using digital archives.



らっしゃいます。

そして、これがネットワーク版になりましたら大変おもしろいのではないかと思います、こういうのやりたいなと思っております。以上です。(拍手)