

領域2

博士論文発信支援
パッケージ開発プロジェクト

東京大学情報システム部情報基盤課学術情報チーム
小松 陽一

概要と背景

東京大学では、学内で生み出される世界水準にある研究成果をインターネット経由で国際的に発信するシステムとして、2006年4月に「東京大学学術機関リポジトリ (UT Repository)」を公開し、サービスを開始しました。リポジトリには、学位論文、学内紀要論文、科研費報告書、COE 報告書、学術雑誌への投稿論文、学会発表論文など、様々な研究成果が電子化されて集中的に蓄積・保存され、学内外に発信されています。

博士論文は、文献調査対象になりますが、一部の商用出版されるものを除いては、学術雑誌のように流通しないため、コンテンツへのアクセスが良くありません。国立国会図書館やリポジトリ等にオンライン上に掲載されている場合もありますが、一般的には閲覧のためには図書館に向く必要があります。これまでの学位規則の運用で、博士論文は冊子で作成の上、国立国会図書館に納本し、大学図書館にも納本するということが行われており、求めれば所蔵場所に赴いて閲覧できるということは確保されているものの、学術利用が極めて困難な状態にあり、電子化による発信・利用環境の改善は、喫緊の課題となっています。そこで、附属図書館及び情報基盤センターでは、大学院生を対象に「東京大学学術機関リポジトリ」の周知を図り、学位論文の電子ファイルのリポジトリへの登録を促進してきました。

本プロジェクトは、これまで東京大学が培ってきた経験を活かし、博士論文を収集・発信するために必要とされる制度化と、デジタルで受け入れて発信する仕組みについて、その具体的な実現手法を他大学と共有できるようにするプロジェクトです。

プロジェクトは国立情報学研究所の学術機関リポジトリ構築連携支援事業の領域2において平成22年度から24年度まで3年間行われ、成果を公開してきました。

内容は、博士論文の収集・発信への準備・制度化に着手してから、具体的な成果発信に至るまでの各段階で必要とされる手法です。具体的には、博士論文登録システム、博士論文検索インターフェース、電子化仕様書のパッケージ化、許諾ノウハウがあります。コンテンツの充実や登録の促進のために、予算やマンパワーが一時的に確保できた場合でも参考になる内容だと思えます。このプロジェクトの全体報告を以下にまとめます。

本報告書を含め、開発パッケージ等の内容は以下のサイトで公開しています。

<http://www.dl.itc.u-tokyo.ac.jp/ut-repository/>

目次

1.	大学内の運用整備（制度化）に関して	1
1-1.	学位論文の提出時期の集中について	1
1-2.	登録手順	1
1-3.	学内広報	2
1-4.	説明資料の作成	3
1-5.	許諾書のカスタマイズ	3
2.	学位論文検索インターフェースの開発・パッケージ化	5
3.	学位論文登録インターフェースの開発・パッケージ化	7
4.	電子化仕様書等の実効性検証	8
5.	課題及び課題解決へ向けての展望	9
6.	遡及登録作業レポート	9
6-1.	作業手順	9
6-2.	著作者リスト作成	9
6-3.	著作者連絡先調査	10
6-4.	メール文面	11
6-5.	メール送信後の反応について	12
6-6.	原本確認調査	14
6-7.	冊子の特徴	14
6-8.	電子化仕様書作成	14
6-9.	見積のための業者への情報提供について	15
6-10.	電子化費用を抑える方法	16
6-11.	リポジトリ登録作業等	16
6-12.	後処理	16

1. 大学内の運用整備（制度化）に関して

1-1. 学位論文の提出時期の集中について
 リポジトリ登録処理は登録者からの申請により開始されます。個人、或いは紀要等の編集者からの登録申請や問い合わせになりますが、それらは年間を通じてランダムに発生することが多いです。一方、学位論文は学位授与手続きとして3月に集中するため（9月頃にも留学生等である程度まとまった数があります）、その流れで登録件数も集中します。他の登録コンテンツでは、紀要等、ある程度の年間における発行スケジュールの影響をうけるものや、年度単位の研究プロジェクトの終了等で年度終了頃に集中しやすいものもありますが、他に比べて学位論文の生産される時期は集中していると言えます。

平成22年度課程博士学位授与数（月別）

授与月	件数
4月	44
5月	14
6月	26
7月	20
8月	0
9月	127
10月	22
11月	12
12月	26
1月	11
2月	19
3月	754
合計	1075

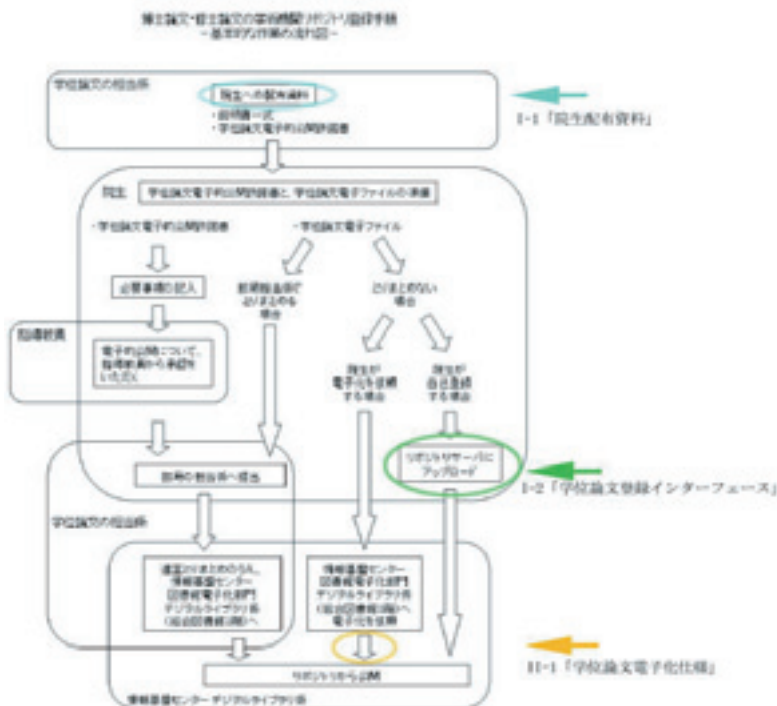
このような状況の中、提出毎に処理するのではなく、とりまとめて集中的に処理することで効率化できます。つまり、博士論文の提出先である部署が3月分を中心にしてそれまでの蓄積分を取りまとめるこ

とで、同様の処理を一度に集中できます。実際には、大学のスケジュールで、3月前後は新学期開始と年度切り替えで、多忙な時期になります。そのため、部局で取りまとめる場合でも、新年度開始後、しばらく後でまとめて提出されることとなります。

1-2. 登録手順

登録手順としては、大まかには2通りあります。部局とりまとめで一括して提出する方法と、個人が登録インターフェースによって登録申請する方法です。

部局とりまとめの場合は、ファイルや許諾の有無に関して部局側で確認してくれるので、提出物の不足が起こりにくく、まとまった数でのやり取りとなります。また、部局での学位授与申請手順の一連の流れに載せて処理している場合はより確実に収集できます。許諾書をカスタマイズして、図書館での公開方法について、全ページ複写を許すかどうかを書かせている部局もあります。部局担当者の作業負担は大きくなります。



個人が登録インターフェースによって登録申請する方法では、部局では、登録インターフェースの入力フォームでファイルのアップロードを行う方法です。許諾書は個人から学内便等で送ってくる方法と、部局で取りまとめて送付する方法があります。前述の方法より部局の負担は少なくなります。学位授与の申請手続きに含めた形で案内することで、より多くの登録が望めますが、手続きに含めずに本人の任意で手続きをさせた場合の申請は少なくなります。

この方法の利点は、登録フォームでメタデータを入力するようにしてあるので、リポジトリの登録担当者の省力化になります。また、部局担当者は案内が中心で実務的な負担が比較的少なく済みます。欠点としてはファ

イルがアップされても、許諾書が来ないなどの確認作業の手間は増える傾向があります。

この2つの代表的な手順以外では、「学位取得者が電子化を依頼する場合」も想定しましたが、実際には、ファイルを持っていることが普通で電子化要求は稀です。また、直接問い合わせにより、メール添付などで受け取る場合もありますが、年間の提出スケジュールから漏れたり（数年後に登録依頼）した場合等でこれも稀です。

1-3. 学内広報

広報として、リポジトリページでの登録コンテンツ募集のページでの説明の他、リポジトリ概要のリーフレットやクリアファイルを配布して継続的に普及に努めています。



より積極的な広報としては、東京大学は2006年からリポジトリの登録を開始し、各部局学務担当者へ説明行脚を行いました。その結果、学位論文の登録に関しては、3部局から積極的な協力を得ました。近年は登録作業時間が不足しているため、継続的に説得することはしていません。紀要類を中心に登録コンテンツが増えつつあり、対応しきれなくなっています。

積極的に関与しない理由としては、本文を

登録しなくとも学位規定を満たしている状況（「基本的に、要旨を電子的に公開し、本文の公開（大学図書館などに納本して、利用者の要望に応じて閲覧出来る状態）はしているのでそれ以上しない。」）で、現状以上のことをする決定的な理由が無いためと思われます。同様の理由で、「様子をみたい」と保留した部局もありました。職員削減の進む中、職員にとっては、新たな仕事になることが障害と思われます。

種 類	部局名	専攻名/タイトル	参加年度等
学位論文 15研究科中 3研究科	工学系研究科 (21専攻中13専攻が参加もしくは参加予定)	機械工学専攻	平成17(2005)年度より *ただし、電気および電子工学専攻は、1995年以降+aについて適及分あり。
		電気機械工学専攻	
		環境化学工学専攻	
		電気工学専攻	
		電子工学専攻	
		化学工学専攻	
		社会基盤工学専攻	
		情報工学専攻	
		航空宇宙工学専攻	
		地球惑星工学専攻	
		ロボティクス工学専攻	
		応用化学工学専攻	
		化学システム工学専攻	
		化学工学専攻	
		化学システム工学専攻	
	平成18(2006)年度より		
	新領域創成科学研究科 (12専攻中 6専攻が参加)	基礎化学専攻	平成17(2005)年度より
		基礎物理学専攻	
		基礎工学専攻	
		基礎生命科学専攻	
		基礎情報科学専攻	
		基礎地球科学専攻	
		基礎宇宙科学専攻	
		基礎システム科学専攻	
	情報理工学系研究科 (6専攻中 3専攻が参加もしくは参加予定)	電子情報工学専攻	平成17(2005)年度より *ただし、1998年以降+aについて適及分あり。
		応用情報工学専攻	
		社会情報工学専攻	
平成18(2006)年度より			

積極的に参加している部局からは、その後参加専攻が増えたり（工学系研究科社会基盤学専攻）、卒業生からも「他の人達があるのに自分の論文が無いので登録したい（新領域創成科学研究科）」等問い合わせがあります。登録が組織的に進んで増えれば増えるほど申請も加速するものと思われれます。逆に、一斉に許諾の問い合わせを行い、一度許諾の回答をした人が後に許諾しなかった人の存在を知って、自分も非公開にしたいと訂正する人もいます。つまり、登録しようとする動機には迎合的な雰囲気があります。

他に、博士論文を

Book Park (www.bookpark.ne.jp) で公開するためリポジトリには登録しないという部局があります。また、潜在的に、人文系の学科では、将来は出版する予定が多く、リポジトリ登録に消極的な考えもあります。

1-4. 説明資料の作成

説明内容は、リポジトリの概要と、登録申請の流れ、許諾書の記述方法について資料を作成し、それらを使った学位論文登録の促進を依頼する内容にしました。

<http://www.dl.itc.u-tokyo.ac.jp/>

ut-repository/ にて公開しています。

- ・教員向依頼
- ・事務方向依頼

- ・ 院生配布資料（部局事務とりまとめの場合）
- ・ 院生配布資料（学位論文登録インターフェース使用の場合）

リポジトリの概要説明に加え、教員向けには、教員の許諾書での公開の可否についての許諾の説明、事務方向の内容では、案内と配布、取りまとめ方法について説明しています。院生向け資料では、部局取りまとめの場合と、インターフェースを用いた自己登録の場合とで資料を別に作成しました。どちらの方法かを決めた上で、部局に渡す資料を選び、部局担当者がそれをカスタマイズして、配布資料とする素案になっています。提出方法等学内で統一出来る場合は、できるだけ統一した様式の方が効率的だと思います。

運用手順を定めて登録申請できるようにした部局は増えてはいますが、実際の申請数はごく少数の部局もあります。次ページは、2種類の方法の選択具合と手順作成部局の一覧です。

1-5. 許諾書のカスタマイズ

必要な項目の入った汎用的な許諾書を用意しましたが、ほとんどの部局で何らかの変更や追加をして使っています。

例えば、許諾確認において、「公開する」

学位論文のリポジトリでの公開手順を定めている研究科

研究科名	許諾書	本文電子ファイルの提出先	備考
医学系研究科	独自タイプ	部局事務窓口へ提出	
人文社会系研究科	独自タイプ	部局事務窓口へ提出	当座非公開
工学系研究科電気系	独自タイプ	リポジトリサーバへアップロード	
工学系研究科(電気系以外)	独自タイプ	リポジトリサーバへアップロード	
理学系研究科	独自タイプ	リポジトリサーバへアップロード	
農学生命科学研究科	独自タイプ	部局事務窓口へ提出	
経済学研究科	独自タイプ	リポジトリサーバへアップロード	
総合文化研究科	独自タイプ	部局事務窓口へ提出	
教育学研究科	独自タイプ	部局事務窓口へ提出	
数理科学研究科	独自タイプ	部局事務窓口へ提出	
新領域創成科学研究科	独自タイプ	部局事務窓口へ提出	
情報理工学系研究科	独自タイプ	リポジトリサーバへアップロード	
情報学環・学際情報学	独自タイプ	部局事務窓口へ提出	

許 諾 書

平成 年 月 日

東京大学附属図書館長 殿

著作者

氏名： _____ 印

所属： _____

(専攻/コースまで記入)

学籍番号： _____

連絡先

Tel: _____

E-mail: _____

種別	<input type="checkbox"/> 博士論文 <input type="checkbox"/> 修士論文
論文題日(和文)	
論文題日(英文/その他)	
論文提出日	平成 年 月 日
電子公開の可否	<p>上記学位論文を、「東京大学学術機関リポジトリへの提供論文利用許諾条件」にしたがって、公開を許諾します。</p> <p>公開時期は、<input type="checkbox"/>審査終了即時 <input type="checkbox"/>期日指定(学位授与日から _____ 月以降)</p> <p>備考： _____</p>
指導教員の確認	<p>上記学位論文について、上記条件での複製、電子公開を認めます。</p> <p>指導教員氏名： _____ 印</p>

職員記入欄	学位番号: 博/修 甲/乙 第 _____ 号	学位授与年月日: 平成 年 月 日
	備考: _____	公開年月日: _____ : 平成 年 月 日

という回答ながら、選択肢に「公開日は著者及び指導教員両方からの指示により公開」というものを追加し、研究室の実験技術漏洩や特許関係の技術漏洩を嚴重に管理するという運用をするところもあります。ちなみに、同様の理由で、公開後に、研究室の教員から、公開を取りやめて欲しい旨連絡が来て公開を中止することもあります。

他には、図書館利用における全ページ複写の可否を予め許諾書で回答させ、閲覧の運用でそれに沿って複写対応する部局や、公開日について、公開時期を審査終了時（学位授与後、データ提出時、即時）、6ヶ月後、1年後、2年後、期間任意指定等多様な選択肢を設ける部局、共同研究者の有無と氏名（医学系研究科で公開許諾の場合に明記）と、各々からの独自の許諾書を添付する部局もあります。

2. 学位論文検索インターフェースの開発・パッケージ化

検索インターフェースを開発し、操作性を向上させました。利用の増加により、登録促進に寄与することが主な狙いです。意見交換会等での意見や、利用者からの要望を参考に改良していきました。以下のような機能があります。

・学位論文特有事項の element の検索

DRF@Okayama のメタデータ標準案を受けて、従来 description にまとめて記述していた学位論文特有事項（例：学位記番号）を個別の element に分け、それらの element を検索するように改良しました。

学位名称：dc.description.degreelevel

例：博士（農学）

学位：dc.description.degreelevel

例：doctoral

報告番号：dc.description.thesisno

例：甲第12345号

学位記番号：dc.description.degreelevel

例：博農第123号

研究科・専攻：

dc.description.degreegraduate

例：Department of Forest Science, Graduate School of Agricultural and Life Sciences

（農学生命科学研究科森林科学専攻）

学位授与年月日：dc.date.issued

例：2005-03-09

学位授与機関：

dc.description.degreegrantor

例：University of Tokyo（東京大学）

学位分野：dc.description.degreediscipline

例：Agriculture（農学）

DRF@Okayama のメタデータ標準案では、「URI 以外の所在情報（冊子等の所蔵情報）」と「出版情報（ISBN 等）」も国内固有項目として挙げられていますが、学位論文必須とも思えないので、また、他の element で記述可能と思われたため外しています。また、学位名称、学位、報告番号、学位授与機関、学位分野については、NDLTD の Union Catalog へのハーベスティングを前提として設けました。そのため英語表記が先に来ています。学位記番号については検索の便を考えて設けました。研究科・専攻については、自分の所属の専攻での登録内容を確認したいという教員よりの要望で設けました。つまり、検索せずに組織のディレクトリを辿って表示させていく方法（DSpace でいうコミュニティ）がありますが、辿ってブラウズしていると見逃す可能性があっても、検索であれば抽出可能になるというものです。なお、address にしなかったのは、論文博士のように所属と審査委員会を設けた研究科が異なるケースがあるためです。学位授与年月日については、data.issued で代用することにしました。

UT Repository Disseartation Search (学位論文検索) simple mode

Options (詳細条件での絞り込み)

Language (論文の言語): not specified (指定なし) English (英語) Japanese (日本語) German (ドイツ語)

Degree (学位の種類): not specified (指定なし) Doctor, course-based (課程博士) Doctor, dissertation-based (論文博士) Doctor (課程博士+論文博士)

Degree (学位の種類): not specified (指定なし) Doctor, course-based (課程博士) Doctor, dissertation-based (論文博士) Doctor (課程博士+論文博士)

Subject (学位の種類):
 Law (法学) Engineering (工学) Agricultural (農学) Economic (経済学)

School/Department (研究科・専攻):
 Graduate School of Medicine (医学系研究科) Aging Sciences (高齢学専攻) Surgical Sciences (外科学専攻) Functional Biology (機能生物学専攻) Health Sciences and Nursing (健康科学・看護学専攻) International Health (国際保健学専攻) Social Medicine (社会医学専攻) Cardiovascular Medicine (循環器内科学専攻) Gastrointestinal Surgery (消化器外科学専攻)

Thesis No. (報告番号):
e.g. #10000 (Numbers such as "10000" is also acceptable. 例 #10000 (番号のみ可))

Degree No. (学位記番号):
e.g. 理工第1234号 (Numbers such as "1234" is also acceptable. 例 理工第1234号 (番号のみ可))

Degree Date (学位授与年月日): Year/年 Month/月 Day/日 ~ Year/年 Month/月 Day/日
e.g. 2008 Year-2009 Year, 2008 Year 4 Month-2010 Year 3 Month, 2008 Year 3 Month 24 Dec-2009 Year 3 Month 24 Dec 例 2008年-2009年, 2008年4月-2010年3月, 2008年3月24日-2009年3月24日

Order by (ソート順): title (タイトル順) degree date (日付順) author (著者名順) thesis No. (報告番号順) degree No. (学位記番号順) descending-order (逆順にする)

Output format (出力形式): Web (HTML) CSV (Excel) Refworks

UT Repository Disseartation Search (学位論文検索) detail mode

Title (論文タイトル):
The title written in the same language as the body of the thesis. 本文の言語が日本語であれば日本語のタイトル、英語ほかの言語であれば当該言語のタイトルも入力してください。

Title alternative (他言語の論文タイトル):
The title written in other languages. 他言語のタイトルで検索する場合は入力してください。

Author (著者名):
The author name written in the same language as the body of the thesis. 本文の言語が日本語であれば日本語の著者名も、英語ほかの言語であれば当該言語の著者名も入力してください。

Author alternative (他言語の著者名):
The author name written in other languages. 他言語の著者名で検索する場合は入力してください。例: 私は半角スペースで区切って入力してください。

Language (論文の言語): not specified (指定なし) English (英語) Japanese (日本語) German (ドイツ語)

Degree (学位の種類): not specified (指定なし) Doctor, course-based (課程博士) Doctor, dissertation-based (論文博士) Doctor (課程博士+論文博士)

Subject (学位の種類):
 Law (法学) Engineering (工学) Agricultural (農学) Economic (経済学)

School/Department (研究科・専攻):
 Graduate School of Medicine (医学系研究科) Aging Sciences (高齢学専攻) Surgical Sciences (外科学専攻) Functional Biology (機能生物学専攻) Health Sciences and Nursing (健康科学・看護学専攻) International Health (国際保健学専攻) Social Medicine (社会医学専攻) Cardiovascular Medicine (循環器内科学専攻) Gastrointestinal Surgery (消化器外科学専攻)

Thesis No. (報告番号):
e.g. #10000 (Numbers such as "10000" is also acceptable. 例 #10000 (番号のみ可))

Degree No. (学位記番号):
e.g. 理工第1234号 (Numbers such as "1234" is also acceptable. 例 理工第1234号 (番号のみ可))

Degree Date (学位授与年月日): Year/年 Month/月 Day/日 ~ Year/年 Month/月 Day/日
e.g. 2008 Year-2009 Year, 2008 Year 4 Month-2010 Year 3 Month, 2008 Year 3 Month 24 Dec-2009 Year 3 Month 24 Dec 例 2008年-2009年, 2008年4月-2010年3月, 2008年3月24日-2009年3月24日

Abstract (論文要旨):
論文要旨中に含まれるキーワードを挿入した場合は入力してください。

Keywords (論文キーワード):

Order by (ソート順): title (タイトル順) degree date (日付順) author (著者名順) thesis No. (報告番号順) degree No. (学位記番号順) descending-order (逆順にする)

Output format (出力形式): Web (HTML) CSV (Excel) Refworks

(参考 URL) <http://drf.lib.hokudai.ac.jp/drf/index.php?DRF=Ookayama>

- ・日本語版・英語版・日英併記版インターフェースの作成

従来は日英併記版のみであったが日本語版と英語版に分けてほしいという要望を受けて作成しました。

- ・インストールが容易です。代表的な OS に対応しています。
- ・検索結果のソート機能
- ・検索のファイル出力機能

これは、文献管理ツール Refworks へのデータインポート用にファイル出力が行える他、CSV 形式での出力もできます。CSV 形式は色々と応用できますが、目的の1つは、部局の教務関係者から、担当部

局で登録済みリストを管理用にデータ抽出して欲しいという要望がたまに来るので、検索インターフェース上で抽出できるように作成しました。

ダウンロードページ：

<http://www.dl.itc.u-tokyo.ac.jp/ut-repository/>

利用例ページ：

http://repository.dl.itc.u-tokyo.ac.jp/dsearch/search_form_bilingual_simple.html

3. 学位論文登録インターフェースの開発・パッケージ化

UT Repositoryへの学位論文の新規登録用フォームです。
投稿する論文に関する情報を下のフィールドへ入力して送信してください。

前に登録した論文を訂正する場合は、更新にチェックを、再度全ての項目を埋めて送信してください。

新規 更新

著者名を入力してください。本文の言語が日本語の場合は日本語の著者名を本文の言語が英語の場合は英語の著者名を入力してください。
Enter the names of the authors of this item below.

姓 (family name) 名 (given name)

著者名 (Author):

他言語の著者名を入力してください。本文言語が英語の場合、著者名に英語の著者名を入力。ここでは日本語の著者名を入力してください。
Enter the names of the alternative authors of this item below.

他言語の著者名 (Authors alternative):

学位論文のタイトルを入力してください。
Enter the main title of this item.

タイトル (Title):

別言語のタイトルがあれば入力してください。
If the item has alternative title, please enter them below.

他言語のタイトル (Title alternative):

研究科と専攻を選択してください。選択肢のなかに存在しない場合は下のボックスに記述してください。

研究科と専攻 (address):

学位授与日を入力してください。
Enter the date of award.

年 月 日

学位授与日 (date issued): 年 月 日

論文本文の言語を選択してください。
Select the language of the full text.

論文本文の言語 (language):

学位の種類を選択してください。
Select the degree type.

学位の種類 ():

学位の種類を入力してください。
Enter the degree type.

学位の種類 (): (M0(博士(工学)) 博士(理学))

学位記番号を入力してください。不明の場合は入力不要です。
Enter the degree number. It is not necessary to enter if unknown.

学位記番号 ():

抄録があるか無いかのみ指示してください。抄録があれば本文から抄録を抽出者が変更します。
Indicate only whether there is an abstract. If there is an abstract, the abstract will be extracted from the full text by the user.

抄録 (abstract): あり なし

アップロードする学位論文本文を選択してください。
Select the full text of the thesis to be uploaded.

本文ファイル (file):

※アップロードするファイルの上限値は10MBとなっています。10MB以上のファイルをアップロードする場合は下記の問い合わせ先へお問い合わせください。
※文字の項目が必須項目です。

連絡先のおあなたのメールアドレスを入力してください。
Enter your contact email address.

メールアドレス (mail_address):

リポジトリの登録では、学位記番号等、学位論文特有の必要項目で、一般的な登録コンテンツには無い属性があります。それらの項目を追加しました、またできるだけ選択式を増やして入力文字列のプレを防いだり、迷わず入力できるよう項目を細分化してあります。入力情報が不足していると、本人との連絡で、確認作業に手間がかかるので、代表的な必須事項については、登録用フォームで入力してもらうことで、代理登録時の入力・確認作業を軽減することが可能になります。この登録用フォームについては、他機関によるテストも行い、操作性の向上等改良を行いました。他に次のような機能や特徴があります。

- ・代表的な OS に対応し、インストールが簡単にできるパッケージにしています。
- ・登録確認メール機能

登録時に入力してあるメールアドレスに、フルテキストファイルやメタデータがきちんとアップロードされたことをユーザにメールで知らせる機能です。登録の確認と記録になります。

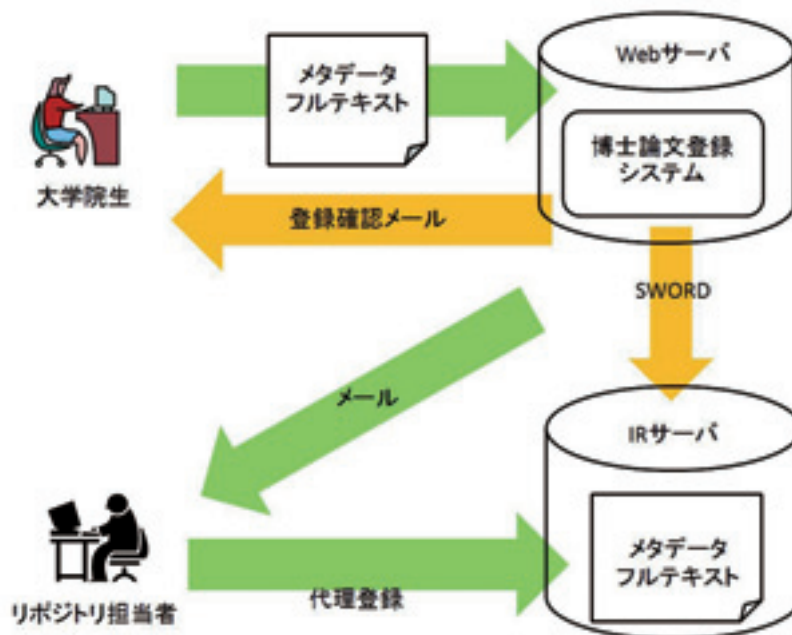
- ・SWORD 形式のファイル出力機能

代表的リポジトリ構築ソフトである DSpace には SWORD 形式によるアイテム登録機能があるため、博士論文登録システムでアップロードされたデータを SWORD 形式にしてファイル出力できるようにしました。

- ・日本語版、英語版、日英併記版の作成

従来は日英併記版のみでしたが日本語版と英語版に分けてほしいという要望を受けて作成しました。

- ・インストールが容易です。代表的な OS に対応しています。



4. 電子化仕様書等の実効性検証

冊子体の電子化及び OCR 処理による透明化テキストの付与という条件で仕様書を作成し、国立国会図書館の電子化事業対象分は除

いた学位論文を対象に、電子化を行いました。OCR 処理による透明化テキストの付加を仕様書とする場合、数学記号や化学記号等は認識が難しく、緻密なチェック作業を行おうとすると費用がかかることが分かり、予算と便益についてどのようにバランスを取るか

という問題がありました。またこの作業の際に著者に対してリポジトリ登録の承認の確認を行いました。その際の、著者の連絡先の探し方等、様々なノウハウを得ることができたため、それらを含めてレポートにまとめました（後述）。

5. 課題及び課題解決へ向けての展望

依頼から登録、また、遡及的な電子化の際の依頼方法等を整理し、検索利用の促進を検討してきました。しかし、新規に生産される学位論文を網羅的に収集蓄積するには程遠い状況です。次のステップとしては、「リポジトリには当たり前のように登録するもの」というコンセンサスを得ることが必要になりますが、現状では決定的な理由がありません。登録実績が増えることで、「周りもやっているから自分も登録しよう（自分の部局も力を入れよう）」という動きは他のコンテンツも含めてリポジトリ全般にあるのですが、学位論文の場合は進展が遅い感じです。

折しも、平成25年4月からの学位規則の一部改正により、博士論文の公表が「印刷」から「インターネット」により公表することになりました。公開は基本的にリポジトリによるため、登録件数は飛躍的に増える予定です。こうなると、これまでの登録促進のための対策よりも、規則改正に関する運用の変更と、増加する作業量へどのように対応するかが課題となります。手続きの簡略化、運用方法の多様性の排除、担当の割当て（分業）をしないと対応しきれないと予想されます。まずは情報収集して対応を協議し、処理対象が大幅に増えることに関しての効率的な運用体制の確立が優先されます。

東京大学は多数の部局から構成され、歴史的に独立性が高いため、全学的な統一運用がし難い面がありますが、学位規則改正をきっかけに学内でのリポジトリの重要性が高まり全体的な登録増加が予想されます。紀要類等

も含めて、継続的にコンテンツが登録されるようになると、それらの対応だけで作業時間が増えていきます。これまで、学位論文に限らず、コンテンツの充実のために、冊子体からデジタル化するということをして来ましたが、現状で生産されるものの登録作業時間が確保できなくなると、遡及的な作業は難しくなっていくと思われます。将来の課題としては、許諾確認に関する根本的な省力化や作業人員の確保が必要になると予想します。

6. 遡及登録作業レポート

冊子体の学位論文から公開用PDFファイルを作成して登録する場合の作業レポートを以下にまとめます。これは、初めからファイルで提出されたものではなく、冊子で提出されたもの、主に、過去分の遡及的登録作業を想定しています。過去分を網羅的に登録することは大変むずかしいのですが、その場合でも効率化できるポイントは何か、また、何が障害になるのか等、実際に作業の参考になるようにレポートにまとめました。なお、部分的に学位論文に限定しないリポジトリ登録全般の内容も含んでいます。

6-1. 作業手順

以下のような手順で作業を実施しました。

- ・著作者リスト作成
- ・著作者連絡先調査
- ・原本確認調査
- ・登録確認依頼
- ・電子化仕様書作成
- ・リポジトリ登録作業等
- ・後処理

6-2. 著作者リスト作成

本学では、以前より論文書誌と要旨を収載した「東京大学学位論文データベース (<http://gakui.dl.itc.u-tokyo.ac.jp/>)」を運用しているので、このデータベースからデータ

を抽出し、著作者リストを作成しました。このようなデータベースが無い場合は、学位授与担当部署からリストをもらう等、リストの元となる情報が必要です。

抽出数に関しては、学位論文の著作者全件分のリストを作成することは可能ですが、量的にその後の作業を完遂できないことは明らかであったため作業の完結できる件数で対象を絞りました。

また、論文博士と課程博士に分けて対象としていきました。論文博士の冊子は担当係の職場がある総合図書館に保管されており、原本確認調査がし易いので、プロジェクト開始年度に着手対象としました。次に、課程博士を対象とした場合には、論文博士の場合と異なり、冊子現物はキャンパス内に散在する各部局図書室にあるため、対象部局を限定しました。農学系研究科を選択したのは、平成22年度での一連の作業の中で、農学系研究科の博論はデータ資料が多いことから、市販の出版物に掲載しづらいこと、就業先が国や地方自治体が多いことから、連絡先が比較的判明しやすいことが予想できたためです。

論文博士、課程博士のいずれの対象にせよ、抽出リストの候補からは、国会図書館の博士論文のインターネット公開分は除外しました。

6-3. 著作者連絡先調査

・許諾確認の連絡方法の選択

連絡方法は、往復はがき等による郵送、電話(FAX)、メール等が考えられます。また、調査する媒体としてはネット上の情報を検索調査するのが一番手軽です。

はがきを想定した場合、所属機関の代表所在の住所は判明しても、どの部署であるとか、支部であるとかの情報までは調べ難く、FAXも同様です。電話も代表番号は判りやすいのですが、所属部署までは判らないことも想定されます、また、同じ説明を繰り返す必要があり、件数が増えると効率がよくあり

ません。メールの場合は、古い情報では届かない場合があるというマイナス面がありますが、対象のメールアドレスが判ればダイレクトに届くので単純であること、送信コストがかからないこと、また、卒業後も研究に従事してメール環境をもっていて、アドレスを公開している場合も想定されるため、メールによる連絡を選択しました。

・メールアドレス調査

遡及対象全体の数からして、実際に許諾を得て、電子化できる数(金額)を加味すると、目標とする件数があったため、1人を判るまで調査するようなことはせず、判らない場合は見切りをつけて、次々と対象を調査していきました。対象のすべての連絡先を調査することを目指した場合は、ネット上の情報には限界があり、調査会社のごとく、関係者に聞き込みをしていくこととなりますがそれはしませんでした。

基本的な調査方法はいたってシンプルに、「著者名 AND e-mail」という検索式でGoogleを検索しました。あとは検索されたサイトを見て、比較的最近に作成・更新されたサイトであることと同姓同名の他人ではないことを確認した上で、メールアドレスを転記して行きました。特に検索対象が企業である場合、異動している可能性が大きいため古い情報には注意しました。また、同姓同名については、学歴や取得学位が掲載されていればそれを確認し、掲載されていなければ研究分野等が博論のテーマと近いかどうかで同定しました。一番確実に判りやすいパターンは、本人のプロフィールをネット上で公開しており、博士論文タイトルや修了年が記載されている場合で、研究者でもとりわけ教員等で多く見られました。

このような調査の結果として、今のところ実際に依頼した際に別人であるという回答は受けていません。

もちろん、メールアドレスがサイトで公開されていないことも多いので、所属が分かれば、所属ドメインとの論理積や、氏名、執筆時所属、論文名の組み合わせ、CiNii等で現在（最終）の大学等研究機関等の所属組織を検索することも有効な方法でした。

それ以外に以下のような傾向や対策がありました。

- ・海外在職の場合は、氏名をローマ字化して検索した。
 - ・病院については、個人のメールアドレスを公開しているところはあまりない。
 - ・薬学系は製薬会社への就業が多く、「著者名 AND 製薬」とやるとそこそこヒットするが、博論のテーマとは別の薬品研究をしていることが多く、同定が困難。また、製薬会社の合併により同定が困難なケースもある。
 - ・農学系は国・地方自治体の農業試験所等への就業が多い。所属ドメインとの論理積が比較的有効。
 - ・所属ドメインは、組織改組が頻繁でなければより有効。メールアドレスがサイトで公開されていなくても、シンポジウムの幹事のメールアドレス等でヒットする可能性がある。
 - ・企業の研究所は改組が多い。
 - ・組織によってはメールアドレスの付け方のルールを設けているケースもあるので、それが分かれば、それで検索してみる。
 - ・氏名のローマ字化は著作者自身の好みや普段使いで定めていることもあるため、うまくマッチしないこともある。その場合は、博論原本や論文データベースで使用方法を確認するか、いくつかのパターンを試してみる。
 - ・同姓同名もよく存在するが、大抵分野が異なるので判別は付く。
 - ・氏名がわかっている場合は、修了年からほぼ同定できる。次に最終ポストの情報を探し、そのポストをヒントに検索する。
- ・本人ではなく、関係部署にメールした場合、本人に転送するように依頼文をつけて送信したが、その組織の経営者（医者や個人事業者等）の場合は回答に至るケースが多いが、そうでない場合は回答が来ない場合が多い。企業の問い合わせフォームなども効果は薄い。
 - ・教員等比較的安定的ポストは時間が経過しても判明しやすい。業績や博士論文をプロフィールとして公開している場合も多く一番簡単に発見できる。
 - ・地方自治体、政府機関研究所等の比較的安定的なポストは時間が経過しても判明しやすい。
 - ・ポストクや特任教員等期間限定身分は3年前後の期限付きで、これを過ぎると、次のポストがない場合にネット上に論文等発表情報も無くなるので辿れなくなる。この意味では修了後3年前後までを境にそれ以前分は難くなる。
 - ・研究室のホームページにOBとして名前が掲載されている場合もあるが、研究室宛に転送願いをメールしても回答は芳しくない。
 - ・留学生が帰国した場合など、ブログなどで自分のアドレスを公開している人もいてメールアドレスも判明して連絡が取れる人がいる。

6-4. メール文面

調査のため、許諾しない場合の理由を回答していただいたこともありましたが、一番単純な依頼内容では以下の様なメール文面で、許諾のために、○をして返信するだけの内容にしました。

- ・通常の依頼メール文例

XXXX 様
東京大学情報基盤課学術情報チームデジタル・ライブラリ担当 XX と申します。

東京大学で生み出された研究成果を、「UT Repository（東京大学学術機関リポジトリ）」で公開しております。

<http://repository.dl.itc.u-tokyo.ac.jp/>

「学術機関リポジトリ」とは、大学等の学術機関で生産された、さまざまな研究成果（学術雑誌掲載論文や学位論文など）を電子的な形態で学内外に公開することを目的とした、ネット上のサーバです。つきましては、「東京大学学術機関リポジトリへの提供論文利用許諾要件」

<http://repository.dl.itc.u-tokyo.ac.jp/document.html>

をご覧ください。下記に学位論文に關しまして、UT Repository に登録し、ネットワーク上で公開することをご承諾くださるようお願いいたします。以下の許諾内容に○を記入(編集)して、メールでご回答ください。

返信先 E-mail: ir-support@lib.u-tokyo.ac.jp

なお、いただきましたご回答およびご連絡先を、他の目的に使用することは一切ございません。ご承諾いただいた場合、東京大学保管の学位論文をスキャニングして電子化する予定です。

東京大学附属図書館長 殿

著作者 氏名：*****

下記の学位論文について、

「東京大学学術機関リポジトリへの提供論文利用許諾要件」にしたがって、電子的公開を許諾します。()

許諾しません。()

...いずれかに○をご記入ください。

論文名 *****
(20XX 年度博士論文)

他の文面のバリエーションとしては、東大所属者への場合は、リポジトリの宣伝も兼ねる内容を追加したり、ファイルを貰いたい場

合の内容を入れたこともありました。

・リポジトリの宣伝

なお、上記以外にも学術雑誌論文、紀要論文、科研費報告書等についてリポジトリ登録を進めております。ご執筆論文等でリポジトリに登録・公開したいとのご希望がありましたら、お気軽にご相談下さい。

・ファイル利用の内容

ご承諾いただいた場合、東京大学総合図書館保管の学位論文をスキャニングして電子化する予定ですが、もし当該論文の電子ファイル（WORD ドキュメント、PDF 等）がお手元にごございましたならば、ご送付いただけますと幸いです。電子化作業の効率のため、ご協力をお願い申し上げます。

6-5. メール送信後の反応について

返事が来る場合には、半数以上は送信直後1日以内で来ます。その後1週間以内に来ないものは、殆ど来ません。回答期限を定めない場合でも殆どが10日以内で来ます。(回答期限は定めるべきです、電子化処理が始まる頃に回答が来ると効率が悪いです)

連絡の中には、「メールを書く暇もないのでこの電話で承諾という回答でよいか?」という電話もありました。個人的によく知る人だったので、そのような手段を選んだと思われれますが、簡単なメールの返事でさえも省きたいという思いがあるものと思われれます。

いずれにせよ、メールを見た瞬間に「許諾しよう」と判断する場合は、実際の許諾の回答の大半になると思われれます。よって、単純な内容と回答方法がよいと思われれます。例えば、書面による記名捺印などは、より回収率を下げる方法と思われれます。また、電話は手軽かもしれませんが、かけ直し、同じ説明の繰り返し、記録が残らないという点でも補助

的な方法以外ではお薦めしません。

- ・ 回答／許諾を得られる割合
調査対象の3, 4件に1件程度のメールアドレスが判明し、調査対象のうち承諾まで得られたのは約1割程度でした。

- ・ 許諾に至った件数（平成24年度の調査時）
759 調査対象数
232 メールアドレス判明送信数
- 86件許諾回答
- 31件許諾しない回答
- 96件回答無
- 19メールアドレス送信失敗
- ・ 回答までの日数

（平成22年度時の状況）

	依頼件数	回答あり	承諾		不可	検討中等
				電子ファイル受領		
8月依頼 ¹⁾	50	32	25	6	7	0
10月依頼 ¹⁾	130	101	78	5	22	1
11月依頼 ¹⁾	61	40	31	0	9	0
11月依頼 ²⁾	213	141	112	29	28	1
合計	454	314	246	40	66	2

¹⁾東大所属教員 ²⁾東大所属教員以外

（平成23年度時の状況）

	依頼件数	回答あり	承諾		不可	検討中等
				電子ファイル受領		
8月依頼 ¹⁾	213	135	107	40	27	1
9月依頼 ¹⁾	101	70	60	2	9	1
10月依頼	5	5	5	0	0	0
合計	319	210	172	42	36	2

¹⁾後日督促分も含む

- ・ 登録不可の理由アンケート（H22, H23年度）
回答には以下の様なものがありました。
- ①既に市販されている、もしくは市販予定
60（人文系に多い）
- ②自分のHP等で既に公開済み 1
- ③状況について様子を見たい 2
- ④賞味期限切れ 1
- ⑤結論が学位論文作成時と異なる 1

以下は（平成23年度8月23日）に行った依頼に対する返信状況です。諾否いずれにせよ、返信される場合は1週間以内に返信されることが多いことを示します。

依頼（件）	1週間以内返信（件）	1週間以後返信（件）	無回答・エラー（件）
25	11	2	12

- ・ 電子ファイルを要望した場合に得られる件数

2005年度ぐらいを目途に、それ以降の対象者に対しては、電子ファイルをお持ちであれば送ってほしい旨も依頼分には書き添えたところ、お応えいただけたものが多かった。対象年が最近のものであれば、要望するのもよいでしょう。

- ⑥非公開を前提に研究（一部技術が関係機関の意向により非公開） 1

- ⑦理由不明 36

この結果は遡及的な登録のものですが、通常の新規登録においては、特に理工系で、「特許情報に絡む内容、最先端の実験方法を公開したくない。」という理由が多く見られ

ます。学位論文を1年経過してから公開する指定も多くあります（数年経過しても許諾されないものも少なくありません）。

人文系では、①の理由で学位論文の公開許諾しないことが多いのですが、学位論文ではなく、紀要類ではリポジトリに積極的に登録されています。部局で発行している紀要や研究雑誌を発表の場として使っており、学生や研究者が多く投稿しています。そして、発刊後は、寄贈による無料配布や、ウェブ公開されているものであるため、もともとリポジトリに登録しやすいものとなっています。さらに、投稿規定をリポジトリに掲載することを前提に変更してもらうことで、リポジトリへの登録が加速しています。

6-6. 原本確認調査

冊子現物で以下の確認をしました。

- ・市販の出版物か否か

論文博士（特に人文社会系）の場合、市販の著書を以て論文とすることが往々にしてあるため。市販されている場合は対象から外しました。

- ・抜き刷りか否か

これも前項と同様で、論文誌に掲載されたものの抜き刷りを以て論文としている場合は対象から外しました。また、参考論文として抜き刷りを付けている場合は、該当部分については電子化の対象から外しました。医学系や薬学系に多く見られます。

- ・ページ数カウント

除外部分もあります。折り込み図も普通にカウントします。また、標題紙がない論文も存在するため、表紙からカウントしますが、見返し等の白紙部分はカウントしません（ノンブルのみ印刷されている場合はカウント）。著作者によっては図表部分についてはノンブルを入れないなどしているケースもあり、注意します。なお、折り込み図の有無やその他特徴的なこと（例えば写真が貼り付けられているとか）は注記しておきます。外注する場

合の見積もりの重要情報になります。

6-7. 冊子の特徴

以下の様な傾向がありました。

- ・数学系は分量が少なく、人社系、建築系、農学系は分量が多い。特に農学系はデータが資料として付けられることが多い。
- ・医学系、薬学系は共著発表の論文を含める場合が多い。抜き刷りが多い。
- ・図表を最後にまとめているものがある。
- ・冊子体そのものの図表が精細さに欠けるものがあり、スキャン方法では改善できないものがある。
- ・両面印刷と、片面印刷がある。
- ・両面刷りでは、裏側が映るものが多い。スキャン結果も同様になる。
- ・保存状態は良い物が多い。
- ・背表紙部分が割れているものや、装丁の表紙がべたつくものが稀にある。
- ・似た装丁が多い(黒地に金色文字タイトル)が統一ではない。
- ・1ページあたりの文字数、行数はものにより大きく異なる。
- ・古い冊子では、写真を別に印刷（現像）したものを糊付け貼り付けたものがある。
- ・ビニールカバーが付いている場合があり、スキャン作業時に手間がかかる。

6-8. 電子化仕様書作成

登録確認が得られた分については、業者に電子化を依頼します。そのための仕様書には、別紙として、対象となる博士論文のユニークなIDとタイトル、ページ数の一覧表を添付し、見積もり合わせを行いました。業者作業期間を凡そ2ヶ月とし、次ページの表のようなスケジュールで一連の作業を行ってきました。

準備作業のところで、ある程度まとまった量をこなしておく、2回目以降の登録確認依頼等がすぐに実施できます。

準備作業 (リスト作成・原本確認)	↔	登録確認依頼	↔	電子化対象論文確定	↔	仕様書作成・見積もり合わせ依頼	↔	見積もり合わせ・発注日	↔	納品日
			一ヶ月		二週間		二週間		二ヶ月	

6-9. 見積もりのための業者への情報提供について

画像スキャニングにおいて、図表等がある場合は、グレースケールの白黒と、カラーのスキャニングを指定しています。特に、グラフや表以外の写真では、色や濃淡に違いがあり、カラースキャンが良いのですが、面倒だからと全てカラースキャンにすると、微妙に色が混じり、モノクロがくっきりした黒にならないので分けることは必要と思われます。

実際のスキャニングでは、カラー、モノクロ、グレースケールの3種類のスキャン方法になりますが、1論文を順次ページでスキャニングするのではなく、スキャン設定は変更せずに、例えば、カラーのみスキャンして、後で、ページ毎に繋ぎあわせて、PDF となります。コストを考える場合、カラーだから高いというのではなく、スキャン対象の管理の手間次第でコストが変化します。つまり、同一のスキャン方法で全てが取れるならそれはコストが低くなりますが、複雑に混ざっている程、後で繋ぎあわせて確認する手間でコストが増加します。また、モノクロ2値のスキャン画像は、OCR 処理をする場合

に、認識精度が高くなるため、それ以外のものの割合が大きいと、誤認識文字の修正手間が増えるので高コストとなります。モノクロ2値の割合が大きければ、低い見積もり額が出しやすいと言えます。

以上より、見積もりにおいて、大まかには対象リストと、それぞれのページ数と論文数があれば見積もり可能ですが、OCR 処理に精度を求めているので、それだけでは見積もりは高めに設定されます。正確に見積もろうとすると、対象分野（化学記号の多さなど）とスキャン方法（白黒2値とそれ以外の割合、混ざり具合）についての情報が必要になります。

これらの情報を、過去のリポジトリで公開済みのものを例示して参照するよう説明もできますが、公開していない分量が少ない場合、或いは公開分とは違う分野等では比率は大きく異なる場合を想定して、数量を明示する仕様になっています。この下調べは案外面倒な作業ではあります。大量の場合は、サンプル調査でよいと思います。また、同じ分野で繰り返す場合は、前回調査の内容を参考情報として使っても良いと思います。

6-10. 電子化費用を抑える方法

何よりコストがかかるのは、OCR（透明テキスト付与）の作業での目視による確認作業です。数式や化学式、日本語処理(旧字体)等は、誤認識するケースが多く、チェックに手間がかかります。

これをソフトウェア的に処理するのみで、検証なし（認識率の精度を仕様に入れない）か、ある程度の専門分野のソフトウェアのチューニング程度にする方法があります。つまり、「ないよりまし」程度の気持ちで、チェックにあまり手間をかけないようにして実施する方法です。その場合の問題点として、誤ったテキストデータが含まれることが多くなります。理工系では、化学式におけるアルファベットの斜体イタリック体に意味があり、厳密に確認していくと、いくらでも誤認識は出てくる可能性があります。クレームが有った場合の対応としては、透明テキストの付与のし直しか、部分的に、透明テキストを編集することが考えられますが、いずれにせよ手間が掛かりそうで、それらのクレーム報告毎に対応すると仕事にならない可能性もあります。透明テキスト無しの（画像的な）PDF ファイルに差替えの方が現実的かもしれません。

もっとコストを抑える方法としては、OCR 処理をしないのも有り得ます。Google からのアクセスのためのインデックスとして、本文中の文字列すべてが対象となるかならないかの差がありますが、多くはタイトルや著者名で検索されることを考えればそれほどものではないとする方法です。

発注時期について、年度会計の宿命かもしれないが、予算確定が遅い場合、また、電子化対象の選定や確認作業を行ってから発注する場合、どうしても、年度後半に発注するこ

とが多くなります。他の文教機関でも同様で、業者は年度後半は大変忙しく、人のアサインが難しくなります。逆に年度前半は余裕があるので、コストを抑える余地があるとのことでした。

初めに予算額が決まっている場合、その範囲内で収めようとしますが、予測のために、業者に大まかな費用を算出してもらっても、最終的には概算予想より低くなるのがよくありました。相見積もりを利用したことが効果的だったと思われます。

6-11. リポジトリ登録作業等

登録確認作業と関連して特筆すべきことは特にありません。

6-12. 後処理

以下の処理を行いました。（年度により行わなかった処理もあります）

- ・「東京大学学位論文データベース」からリポジトリの当該論文へのリンクを作成しました。
- ・許諾していただいた著者に対して、リポジトリに登録したことそのアクセス URI（著者がリンクをはりたくなった時のために）等をメールでお知らせしました。このお知らせを登録直後にすると、JAIRO 等ハーベスティングされるデータベースに関して、「反映していない」と問い合わせが来ることがあるので、反映（2週間程後）まで待つて報告するのが良いかもしれません。
- ・電子化をこちらで行った場合には、電子ファイルが入用かどうかを確認し、入用であればセキュリティを外した電子ファイルを送付しました。