

分担研究報告書

UMINCTR の検索の工夫と課題

研究分担者 木内 貴弘

東京大学大学院医学系研究科 公共健康医学専攻 医療コミュニケーション学分野

東京大学医学部附属病院 大学病院医療情報ネットワーク(UMIN)センター 教授

研究要旨：

目的： 日本国内の治験・臨床データベースの検索機能を充実させていく為に、UMINCTR をレジストリ運用者の立場から利用者からの要望を整理し、課題と対応策を検討することを目的とする。

方法： UMINCTR に関する検索機能に関する工夫、検索機能上に関する課題を中心とした今後の課題について内容の整理と検討を行った。

結果： UMIN 臨床試験登録システムでは、運用当初に検索システム開始して以降、改良を続けてきた。検索機能の状況については、検索項目の変更、条件として指定可能な項目を見直し、検索結果として取得したい試験情報をより指定し易いよう変更したり、フリーワード検索機能強化が挙げられる。さらに、通常検索機能として、主要な項目については、文字型の項目は文字列で、カテゴリ型の項目はカテゴリの指定で検索が可能である。また、フルテキスト検索機能として、登録されたデータ全体に対して、フルテキスト検索機能を提供している。これによって、より精密な検索を必要としている研究者等に対応している。課題とその対応については、2022年12月末に UMIN はリプレースを実施したが、この際にデータベースソフトである Oracle を最新バージョンに更新した。Oracle の最新バージョンへの更新により、データベースに保存される文字情報の文字コードが UTF-8 となり、仕様変更が必要となった結果、2回検索を実行することで対応をしている。現時点において、上記の機能変更は反映されていないが、近日中にリリースの予定となっている。

結論： UMIN 臨床試験登録システムでは検索システムの提供開始後、改良を続けている。検索機能については、検索項目の変更、条件として指定可能な項目の見直し、フリーワード検索機能強化が挙げられる。さらに、通常検索機能として、主要な項目については、文字型の項目は文字列で、カテゴリ型の項目はカテゴリの指定で検索が可能であり、フルテキスト検索機能も実装している。課題は、2022年12月末のリプレースの際にデータベースソフトである Oracle 更新により、仕様変更が必要となった。現時点では2回検索を実行することで対応をしており、近日中に最新のシステムがリリースの予定である。

A. 研究目的

2017年の臨床研究法の施行に伴い、2018年4月からjRCTが稼働し、先般、治験の新規登録は中止となり、jRCTに新規登録されるようになった。

さらに、2021年のデータヘルス委員会や経済財政運営と改革の基本方針2021（骨太方針2021）では、患者の治験情報アクセス向上のため臨床試験登録レジストリ等、臨床研究に係るデータベースの充実の推進（網羅性や検索性の向上）が必要とされたことに伴い、本研究では、UMINCTRに関して、検索性の向上の取り組みと課題について整理することを目的とした。

B. 研究方法

本研究では、次の内容について検討を行った。

1. 検索機能に関する工夫
2. 検索機能上に関する課題
(倫理面への配慮)

本研究は個人のデータを扱っていないことから倫理面への配慮は必要ない。

C. 結果及び考察

UMIN CTR の使い方を画面の図で示す。図1はUMIN CTR のトップページである。トップページで、「臨床試験の検索」をクリックすることによって、図2の試験情報の検索画面が表示される。図2から、臨床試験情報の検索が可能である。トップページから、「登録・公開された全臨床試験の一覧」をクリックすることによって、図3の試験情報の一覧を表示できる。これは、図2の試験情報の検索を行った結果と同様の形式となっている。各臨床試験情報で「閲覧」をクリックすると、個別の臨床試験登録情報の一覧閲覧表示がなされる。「履歴」をクリックするとUMIN CTR 更新情報の履歴が表示される。これによって、登録後になされたすべての更新の内容を確認することができる。

1. 検索機能の状況

UMIN 臨床試験登録システムでは、運用当初から検索システムの提供を行ってきた。以降、改良を続けており、下記の改良を行ってきた。

・検索項目の変更

検索条件として指定可能な項目を見直し、検索結果として取得したい試験情報をより指定し易いよう変更した。

・フリーワード検索機能強化

検索条件として指定したワードが含まれる試験情報を取得するための機能は検索条件で指定可能な項目の1つとして搭載されていたが、よりの確に取得したい試験情報を取得できるようフリーワード検索をより特化した別機能とした。フリーワード検索したいワードを複数指定できるよう変更し、複数のワードを指定した場合、一致条件を「全てを含む (AND 検索)」、「いずれかを含む (OR 検索)」から指定可能とした。また、フリーワード検索で指定したワードが含まれる項目を1つの項目から全項目まで任意に選択できるよう変更した。

・表示ページ選択方法の変更

UMIN CTR/ICDR に登録された試験情報が非常に増え、指定された検索条件によっては検索結果として取得された試験情報の件数が数万件となる場合がある。特定のページを表示し易くするため、先頭ページ/最終ページを表示するためのリンク、任意のページを表示するためのリンク、1ページ単位で表示ページを変更するためのリンク機能を追加した。

・検索結果一覧変更

試験情報の検索を行い表示される検索結果一覧が見やすくなるよう下記を変更した。

- 表示項目、及びレイアウトの変更。

- 試験情報の試験進捗状況、及び試験情報の更新状況により背景色を変更。

現在、検索のユーザインターフェイスやレスポンスについて、改良の余地がないわけではないが、検索自体の機能としては十分なものと考えている。

1) 通常検索機能

主要な項目については、文字型の項目は文字列で、カテゴリ型の項目はカテゴリの指定で検索が可能である（図1）。検索条件として指定可能な項目は下記の通りである。なお、文字型の項目は日本語項目、英語項目の両方に対して検索が行われる。

UMIN 試験 ID、試験名（試験名、試験簡略名、科学的試験名、科学的試験略称）、試験実施地域、対象疾患名、疾患区分1、疾患区分2、目的1、目的2、その他詳細、試験のフェーズ、主要アウトカム評価項目、試験の種類、基本デザイン、群数、介入の目的、介入の種類、年齢（下限）、年齢（上限）、性別、目標参加者数、実施責任組織機関名、研究費提供組織機関名、他機関から発行された試験ID、試験進捗状況

2) フルテキスト検索機能

登録されたデータ全体に対して、フルテキスト検索機能を提供している。これによって、より精密な検索を必要としている研究者等に対応している（図2）。元々フルテキスト検索は通常検索機能の一項目にすぎなかったが、機能が強化され現在のフルテキスト検索機能となった。

2. 課題とその対応

2022年12月末にUMINはリプレースを実施したが、この際にデータベースソフトであるOracleを最新バージョンに更新した。Oracleの最新バージョンへの更新により、データベースに保存される文字情報の文字コードがUTF-8となった。

この変更により仕様の変更を余儀なくされた。

具体的には、Oracleにおいて文字型項目に指定できるデータサイズの最大値は4000バイトである。臨床試験登録システムの主要アウトカム評価項目や選択基準等の項目のデータサイズは4000であり、そのデータ領域をほぼ使い切っている試験情報も登録されている。登録されている内容をUTF-8に変更した場合、データ量が1.5倍となってしまうため、Oracleにおいて文字型項目に指定できるデータサイズの最大値である4000を超えてしまう。そのため、それらの項目のデータ型を文字型(char)からロングテキスト型(CLOB)に変更した。その変更の影響で検索速度が大幅に低下してしまった。数件程度の試験情報の取得ならとにかく、大量の試験情報を取得する場合、体感としてはっきりと検索速度の低下が感じられるぐらいの速度低下である。

その対応として、検索速度を向上させるための変更を現在行っている。古いバージョンのOracleではサポートされていないが、検索を行った場合に検索結果が何万件であれ、指定した件数に達した場合にそこで検索を終了させるための指定が最新版のOracleではサポートされている。ただし、その機能を使用した場合、検索結果として取得される試験情報の数が検索条件に一致した試験情報の件数ではなく、取得された試験情報の件数となってしまう（100件までを取得するよう指定した場合は取得される件数は100件となる）。その対応として、検索条件に一致する試験情報の件数のみを取得するための検索を実行し、件数のみを取得する。件数を取得後に一覧表示する試験情報のデータを取得するための検索を実行する。2回検索を実行することにより、検索にかかる時間がより長くなるように思えるかもしれないが、実際にこの方法の

方が、特に検索結果として取得される試験情報の件数が多くなるほどはるかに高速化される。

現時点において、上記の機能変更は反映されてはいないが、近日中にリリースの予定となっている。

E. 結論

UMIN 臨床試験登録システムでは、運用当初から検索システムの提供開始後、改良を続けている。検索機能の状況については、検索項目の変更検索として、条件として指定可能な項目の見直し、フリーワード検索機能強化が挙げられる。さらに、通常検索機能として、主要な項目については、文字型の項目は文字列で、カテゴリ型の項目はカテゴリの指定で検索が可能であり、フルテキスト検索機能も実装している。

課題については、2022 年 12 月末のリプレースの際にデータベースソフトである Oracle 更新により、仕様変更が必要となった。現時点では 2 回検索を実行することで対応をしており、近日中に最新のシステムがリリースの予定である。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表
 2. 学会発表
- なし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得

特になし

2. 実用新案登録

特になし

3. その他

特になし

臨床試験登録 center6.umin.ac.jp/cgi-open-bin/ctr/index.cgi?function=02

検索

ソート条件
ソート項目: 登録日昇降/Date of registration DESC

検索条件	
UMIN試験ID	<input type="text"/>
試験名	<input type="text"/>
試験実施地域	<input type="checkbox"/> 複数選択可能 <input type="checkbox"/> 日本/Japan <input type="checkbox"/> アジア (日本以外) /Asia(except Japan) <input type="checkbox"/> 北米/North America <input type="checkbox"/> 南米/South America <input type="checkbox"/> オセアニア/Australia <input type="checkbox"/> 欧州/Europe <input type="checkbox"/> アフリカ/Africa
対象疾患名	<input type="text"/>
疾患区分1	<p>内科</p> <input type="checkbox"/> 内科学一般/Medicine in general <input type="checkbox"/> 消化器内科学 (消化管) /Gastroenterology <input type="checkbox"/> 消化器内科学 (肝・胆・膵) /Hepato-biliary-pancreatic medicine <input type="checkbox"/> 循環器内科学/Cardiology <input type="checkbox"/> 呼吸器内科学/Pneumology <input type="checkbox"/> 内分泌・代謝病内科学/Endocrinology and Metabolism <input type="checkbox"/> 血液・腫瘍内科学/Hematology and clinical oncology <input type="checkbox"/> 腎臓内科学/Nephrology <input type="checkbox"/> 神経内科学/Neurology <input type="checkbox"/> 膠原病・アレルギー内科学/Clinical immunology <input type="checkbox"/> 心臓内科学/Psychosomatic Internal Medicine <input type="checkbox"/> 感染症内科学/Infectious disease <input type="checkbox"/> 老年内科学/Geriatrics <p>外科</p> <input type="checkbox"/> 外科学一般/Surgery in general <input type="checkbox"/> 消化器外科(消化管) /Gastrointestinal surgery <input type="checkbox"/> 消化器外科 (肝・胆・膵) /Hepato-biliary-pancreatic surgery <input type="checkbox"/> 血管外科/Vascular surgery <input type="checkbox"/> 呼吸器外科/Chest surgery <input type="checkbox"/> 内分泌外科/Endocrine surgery <input type="checkbox"/> 乳腺外科/Breast surgery <p>その他の診療科</p> <input type="checkbox"/> 産婦人科学/Obstetrics and gynecology <input type="checkbox"/> 小児科学/Pediatrics <input type="checkbox"/> 眼科学/Ophthalmology <input type="checkbox"/> 皮膚科学/Dermatology <input type="checkbox"/> 精神神経科学/Psychiatry <input type="checkbox"/> 耳鼻咽喉科学/Oto-rhino-laryngology <input type="checkbox"/> 整形外科/Orthopedics <input type="checkbox"/> 泌尿器科学/Urology <input type="checkbox"/> 放射線医学/Radiology <input type="checkbox"/> 麻酔科学/Anesthesiology <input type="checkbox"/> 口腔外科学/Oral surgery <input type="checkbox"/> 脳神経外科/Neurosurgery <input type="checkbox"/> 心臓血管外科/Cardiovascular surgery <input type="checkbox"/> 形成外科/Plastic surgery <input type="checkbox"/> 美容外科学/Aesthetic surgery <input type="checkbox"/> 検査医学/Laboratory medicine <input type="checkbox"/> 手術医学/Operative medicine <input type="checkbox"/> 救急医学/Emergency medicine <input type="checkbox"/> 輸血医学/Blood transfusion <input type="checkbox"/> 集中治療医学/Intensive care medicine <input type="checkbox"/> リハビリテーション医学/Rehabilitation medicine <p>歯学・看護学</p> <input type="checkbox"/> 歯学/Dental medicine <input type="checkbox"/> 看護学/Nursing <p>該当せず</p> <input type="checkbox"/> 該当せず/Not applicable <p>健常人</p> <input type="checkbox"/> 成人/Adult <input type="checkbox"/> 小児/Child
疾患区分2	[未選択/Not selected]
目的1	<input type="text"/>
目的2	[未選択/Not selected]
目的2 - その他詳細	<input type="text"/>
試験のフェーズ	[未選択/Not selected]
主要アウトカム評価項目	<input type="text"/>
試験の種類	[未選択/Not selected]
基本デザイン	[未選択/Not selected]
群数	<input type="text"/>
介入の目的	[未選択/Not selected]
介入の種類	<input type="checkbox"/> 医薬品/Medicine <input type="checkbox"/> ワクチン/Vaccine <input type="checkbox"/> 遺伝子/Gene <input type="checkbox"/> 食品/Food <input type="checkbox"/> 医療器具・機器/Device,equipment <input type="checkbox"/> 行動・習慣/Behavior,custom <input type="checkbox"/> 手技/Maneuver <input type="checkbox"/> その他/Other
年齢 (下限)	数値 単位 不等号

図 1. 通常検索画面

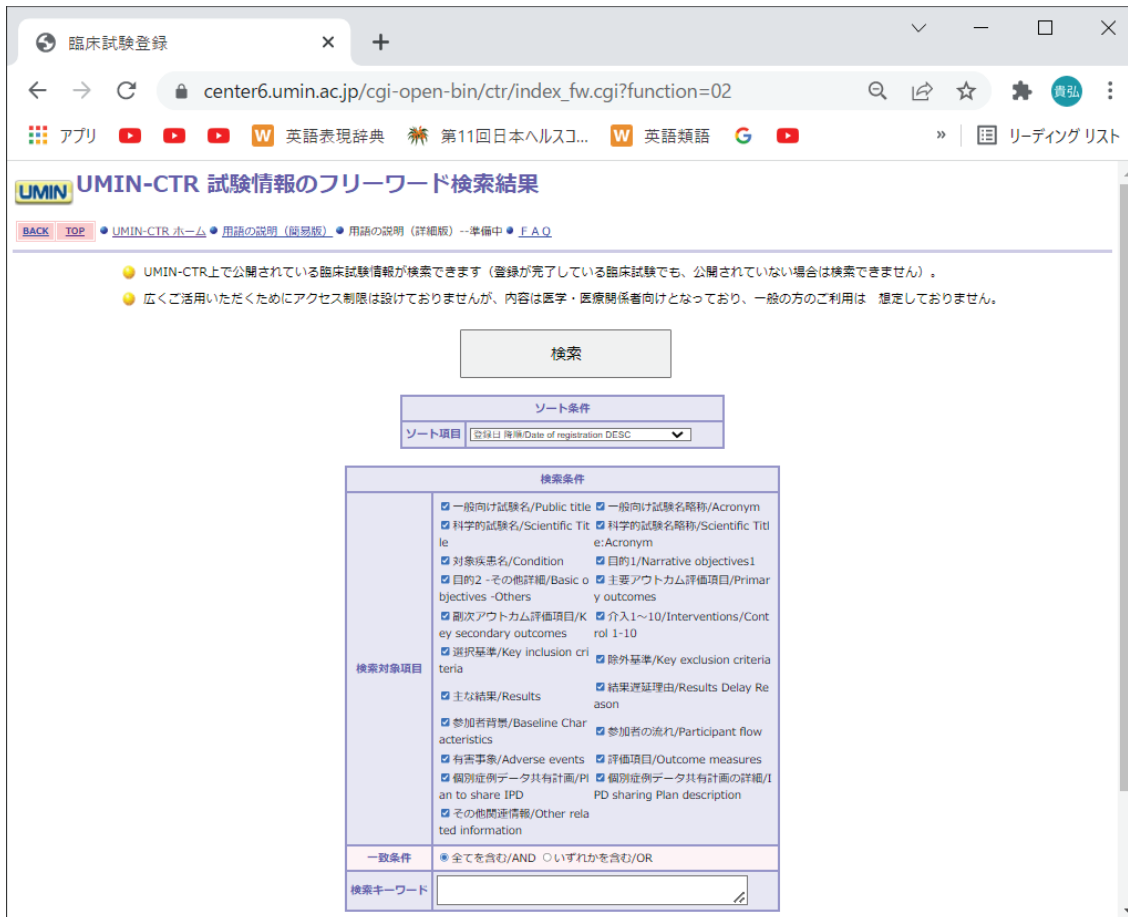


図2. フルテキスト検索画面