ACADE見IC 学術情報メディアセンター／情報学研究科

森 信介 教授

はじめに

今回のACADE見ICでは、学術情報メディアセンター／情報学研究科の教授で2024年４月からは同センター長を務められている森信介先生にお話をいただきました。建物の前はよく通るけど、中でどんなことをしているのかはあまり知らない「学術情報メディアセンター」についてや、森先生が現在取り組まれている「実は身近にある基礎技術の開発」について詳しく伺いました。　（ぷらむ）

ご略歴

京都大学大学院工学研究科電子通信工学専攻博士後期課程を修了後、1998年日本IBMに入社。2007年より京都大学学術情報メディアセンター／情報学研究科准教授。2016年より現職。2024年学術情報メディアセンター長着任。

森先生が率いる「大規模テキストアーカイブ研究分野」の研究室HP

① 学術情報メディアセンターについて

――まず、森先生がセンター長を務められている、学術情報メディアセンターについて教えてください。

学術情報メディアセンター（以下メディセン）は主に情報処理教育センターと大型計算機センターというのがくっついた組織で、建物は南北に分かれています。学生の皆さんが「メディセン」と聞いたときにイメージするのは、南館の方だと思います。１階には学生さんの対応などをする事務が、２階と３階にはパソコンを使う講義が行われるマルチメディア講義室が、４階には研究室が２つあります。メディセン北館にはスーパーコンピューターや情報環境機構、その隣の総合研究５号館には研究室が６つあります。

今の学生はみんな自分のパソコンを持っていますよね。私たちが学生の頃はほとんど誰も持ってないし使うこともなかったのですが、90年代後半ぐらいから使われるようになりました。2000年ぐらいのITバブルあたりからは大学でも使いたいけれど、持っていない学生もいるという状況だったので、メディセンの南館や図書館に来て、共有のパソコンを使うということになっていました。

もともと学内の情報環境の運営もメディセンがやっていたのですが、情報環境機構ができてからはそちらに移管することになりました。移管する前、例えば私は10年ほど前に全学メールを設計しています。学生用のKUMOIは別の方にお願いしましたが、教員用のKUMailは私がやりました。ちなみにKUMailは「くまいる」と発音します（笑）。設計するときに少しこだわったのは、メールアドレスが実名になるようにしたところです。しかも姓・名の順で。文化的に日本人名は先にファミリーネームを言うので、アルファベットで書くメールアドレスでもこの方がいいと思いました。そして姓名の後ろに何かついていますよね。あの文字も全員につけるかどうかは難しい判断だったのですが、ついてない人がいるようにすると、間違いメールが発生しやすくなります。それは私がIBMの研究所に勤めていたときに経験していたので、実はそういう経験が生きて全員に２桁か３桁の文字をつけるという風にしました。

▲IBMの新入社員研修で訪れた野洲事業所前で記念撮影する森先生

――学術情報メディアセンターにはどのような研究室があるのでしょうか。

南館には、教育情報学の研究室と分散システムの研究室があります。北側の総合研究５号館には、まず４階にスーパーコンピューターの研究室とネットワークの研究室があります。そして３階にテキスト関連の私の研究室と、人間計測を含む画像処理の研究室があります。２階には農業統計の研究室と情報教育の研究室があります。総じて言うと学術情報をコンピューターで扱うということになるでしょう。

学術情報の範囲ってすごく広くて、我々の定義でいうと、シミュレーションでやった計算結果も含まれるし、私の研究室が対象にしているような古文書とかの学術歴史資料も学術情報になるでしょう。化学実験のビデオを撮って、そのビデオをAI処理するという研究をするのであれば、そのビデオも学術情報になる。すなわち、学術情報は大学での活動一般だと思ってもらうといいかもしれません。それらの活動から生まれる情報を対象に、コンピューターを使って何かしら知見を得たり便利にする研究ができればいいと思っています。


▲学術情報メディアセンター南館


▲学術情報メディアセンター北館


▲総合研究５号館

はみだしすてーじ

ACADE見ICの記事には、インタビューされた先生の研究室HPをQRコードで載せてはいかがでしょうか。　（他・教　ふぁれすと）
⇒ご提案ありがとうございます！　今回の記事から載せてみました。（読者カードでのご意見・ご感想をお待ちしております！；編）

② 研究について

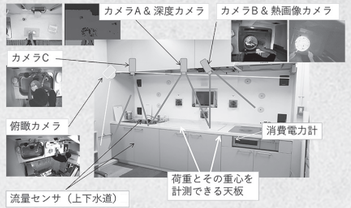
抽象化がなされているから面白くてかつ難しい

――森先生がされている研究について教えてください。

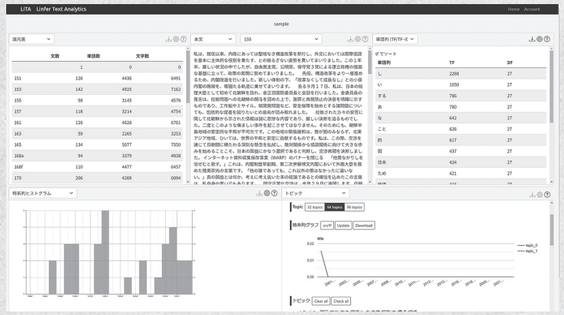
私の研究室はテキストに関係することや言語処理の基礎技術です。皆さんのスマホやパソコンで使われているだろう統計的手法による仮名漢字変換もその１つで、実は身近にあります。その逆の読み推定も研究対象で、この記事を「あかでみっく」と発音するには、少し読み推定をする必要があります。現在力を入れているのは、テキスト中の時間表現と空間表現と人名の絶対値推定ですね。特に人文情報学への応用を考えると、暦が王朝によって異なっていたり、昔は季節で長さが変動する日中を等分した時間だったりして、絶対時間の推定は容易ではありません。また、場所などの空間表現は同じ地名が多数あり、しばしば曖昧です。Florenceはイタリアのフィレンツェですが、実はアメリカにもあります。絶対値は緯度経度で、それを推定します。応用としては、例えば日本では大昔の地震の記録が各地の蔵から出てくるのですが、それらの多数の記録から絶対時間と緯度経度を推定すると、地震の規模や震源を推定できるようになります。そして人名も当然ながら同姓同名があるのでデータベースに紐付けるようにすると、このとき歴史上の人物がここにいたことになっているけど本当か？　この速さで移動できないんじゃないか？　そういう疑問から新事実が発見できる可能性があります。人工知能の推定は100%ということはないので、誤りがいくらかあってもそれを含めて推論するのが難しいところです。それでも精度が高い方がいいので、正解がわかっているものをコンピューターに推定させてみてどれだけ正解できたかを見る、ということを繰り返す「教師あり学習」と呼ばれる方法で、推定の精度を上げていきます。これは皆さんが模擬試験を受けたのと一緒ですね。答えがわかっている模試を受けて「はい80点です」ってなって、また勉強して点数が上がるようなものです。

最近は国立歴史民俗博物館と共同で、埴輪の見た目と言語表現の対応を自動で推定させる研究をやっています。例えば船形埴輪って今の船とは形が大きく違うんですが、人間からすると船に見えなくはないんですね。そういうのを形や模様などの条件を指定することで検索ができるようにしようという感じです。

私の研究室ではテキスト分析ツールの運用もやっています。1,000ページくらいの人間が読める量のテキストでも「ある単語が何回出てきたか教えてくれ」と言われると、もう１回最初から読まないといけないので人間業じゃないですよね。そういうことが簡単にできるツールを作っています。



▲観測装置を備えた特殊キッチン



◀テキスト分析ツール

▼スマホでテキスト分析ツールを体験することができます！

他には、手順書の理解という研究をやっています。簡単に言うと、ロボットがレシピを見て料理を作れるようにしようという研究です。実はメディセン南館の４階にキッチンがありまして、カメラで調理の様子を撮影しながら、重力センサーがついたまな板で包丁の力のかかり具合を測ったり、IHの電力消費量から火加減を測定したりできるようになっています。これらを使って、まずロボットにも理解できるようなレシピを作ろうというところから始めました。そのレシピに何を書いて何を書かないかっていうのが結構難しいんですね。例えば普通の肉じゃがのレシピに「包丁を持ちます」とか書いてないですよね。でもロボットはじゃがいもを切るためには包丁を持たないといけない。ここである種の抽象化がなされているので、これが面白くてかつ難しい。ロボットにとって適切な表現を考えるために、いろんな人に肉じゃがを作ってもらいました。料理があまり上手くない人もいましたが、失敗も大事なデータですからね。

はみだしすてーじ

連休に旅行しないのは損でしょうか？
⇒時間がとれるうちに行っときましょう！　（「もっと旅行しといたらよかった……」という院生の嘆きをよく耳にします。；編）　（農・３　鮒）

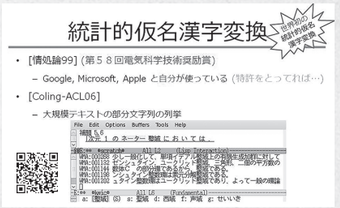
――現在の研究をされるようになった経緯を教えてください。

小学５年生のときにおばあちゃんに貰ったお金でパソコンを買って、ゲームのプログラムを作ったり改造したりしたんですね。当時は１学年160人のうちパソコンを持っているのは５人くらいで、ネットワークも繋がってないので１人でやっている感じでした。ゲームプログラムが載っている雑誌があって、毎月それを紙からタイプして写すんです。それで写し間違えるとゲームがおかしくなるんですよ。それが学ぶきっかけですね。わざとおかしくすると、例えばシューティングゲームだと敵の弾に当たってもやられないようにできるわけですね。弾に当たった判定の文を無効にして無敵にしてみたりしました。すると、非常につまらないゲームになります。でも、そういうのが面白くてプログラムを自分で勉強するようになりました。手先で何か作ったりするのも好きだったので、なんとなく京大の電気系に進学しました。そこに機械翻訳とかをやっていた長尾先生という方がいらっしゃって、言語に対する興味があったというのもあり、その研究室に希望して配属されました。２年生からは第三外国語でフランス語を勉強し始めていて、博士課程の最後の年ぐらいに準１級を取るくらいになっているので、言語処理は性に合っていたんでしょうね。

――研究をしていてやりがいを感じるのはどのようなときですか。

何かしら使えるものを作れると非常にいいですね。私は仮名漢字変換を統計的な手法でするというのを1998年ぐらいに論文で書いています。それまでルールベースで変換していたのを確率モデルで機械学習的に全部バシッとやってしまうという手法で、この技術は多分皆さんも使っていますね。そういう使われるものの基礎を作っていると非常にやりがいがあります。だから、製品に関われる企業に行くのもいい経験になると思って、博士を出てからIBMの研究所に行きました。企業だと「数年のうちに製品になるような技術を研究してください」みたいな感じで、少し短い期間での実用を目指すような研究テーマを求められます。何でもいいってわけじゃないから、そこがまあ難しい。それに対して、大学はある意味何でもいいと言えばいいですね。論文を書いて長い目で人類に貢献すればいいという価値観なので、その辺は違います。

そういうことで、研究するという意味では大学でも企業でもいいんですけど、企業の方は製品に関われるとか給料がいいとかのメリットがある一方、ずっとやりたい研究をするというわけにはいかないでしょう。大学では研究テーマを自分で決められるところが本当にいいですね。



▲実は皆使っている統計的仮名漢字変換の技術

③ 京大生へ

大学は定額学びたい放題

結局論文は英語で読み書きするので、英語の勉強は大事です。私はフランス語の勉強をかなりやっていたことが英語を勉強するときに生きましたが、それでも英語で満足に書けるようになるまでは結構大変でした。あと、読書も大切です。人文情報学をやっていると、本当によく物事を知っている先生と出会うので、そういう人と話すときにやっぱり色々と知っておくのは大事かなと思います。授業はもうちょっと出とけばよかったかな。当時の私は「勉強は教科書を読めばできるからいいや」と思っていたんですが、授業では意外と与太話が面白いはずなんです。教科書に書いていない、先生の感覚や体験談みたいなのを聞けたりする機会として、実は大事だったのかなと今になって思います。

大学は定額学びたい放題なので、結局はやりたいことをやっておくのがいいんじゃないですかね。「教員も含めた大学設備をフル活用しましょう」ということです。私は修士課程のときも、文学部のフランス語の授業に行っていました（笑）。英語以外の言語の授業は学生が少ないからオススメです。自分の幅を広げられますし、国際学会なんかでその開催地の言語が話せたりそこの歴史や文化を知っていたりするとより面白いですから。さっき言ったような、教養のある人と話すときにも役立ちます。

私はたまたま初めに専門にしたものがしっくりきましたが、修士から専門を変える人も多々います。やりたいことを余分に２つか３つやってみて、本当に面白いと思うのか自分で試してみるのがいいと思います。また、情報系では博士課程の後も企業に行くこともできるので、さらに進学するのも悪くないですね。

――貴重なお話をいただきありがとうございました。

はみだしすてーじ

11月号には暑すぎと書いてありますが、逆に寒すぎです
⇒去年の秋はあっという間でしたね。まだまだ寒い日が続くので体調管理に気をつけましょう！　（総・１　p型半導体）
（春が待ち遠しい……；編）