

世界の Blogosphere をつなぐ共同人力翻訳プラットフォーム OpenLingual

同志社大学 工学部 情報システムデザイン学科 前山 晋哉
2007年6月29日

目次

1. 背景
2. 提案内容
3. プレゼンテーション
4. 本プロジェクトの意義
5. 開発の方法、スケジュールおよび予算
6. 過去の成果物
7. 提案者について+将来のソフトウェア技術について

1. 背景

現状 - Blogosphere の隔絶

昨今、インターネットではブログの存在が一般的になってきている。その数は数千万、あるいは一億を超えるとも言われ、毎日 10 万を超える恐ろしいペースで新しいブログが誕生している¹。中には多くの人にとって価値ある情報を配信するものもあり、そういったものは確実にインターネットユーザーの人気を獲得していると言える。

ブログはインターネット上の気になる記事について論評する（Web を Log する）ことからその名（Blog）を得たこともあり、リンク・トラックバックなどにより他のブログや Web 上リソースへ多くの結びつきを持っている。こうしたブログ間の結びつきが形成する「ブログ界」を Blogosphere とよぶ。

Blogosphere のイメージは、複数のブログが形成する小宇宙のようなものと考えてもらいたい。ここで重要なのは、インターネット全体がひとつの大きな Blogosphere と重なっているわけではなく、インターネット上に複数の Blogosphere が存在することである。

いくつもの Blogosphere 間を隔てるのは、ひとつには当然、テーマやジャンルだ。手軽に作れるオリジナルレシピを紹介する料理ブログと、最新の携帯電話事情を紹介するニュースブログとが別々の Blogosphere にあるのは納得していただけるだろう。これは情報のまとまりを作る意味で有意義な壁と言える。

問題はもうひとつの壁である。それは、言語だ。

たとえば手品についてのブログを考えてみよう。日本語で手品に関する記事を定期的に更新しているようなブログは多数存在するだろうし、それらの間にはおそらくリンクやトラックバックのネットワークがあり、ひとつの Blogosphere を形成している。一方、手品の本場はアメリカ、ラスベガスだ。アメリカで活躍するマジシャンたちの中にもブログを持つ者がいて、それらのブログはひとつの Blogosphere を形成しているはずだ。さて、この二つの Blogosphere は本質的に同じ情報のまとまりなのだが、英語と日本語と言う壁があるせいで、決定的に隔絶されてしまっている。

すなわち、日本語の手品ブログは英語のマジックブログに言及することは稀であり、仮にしたとしても勇敢に英語の記事に立ち向かう読者はごく少数派である。そして逆もまた然りだ²。手品に限らず、日本語と英語に限らず、本来ひとつであるべきさまざまな Blogosphere が言語の刃によって分断されてしまっているのが現状なのである。

理想 - Blogosphere をつなぐ

そこで、同じテーマならひとつの Blogosphere に結合してしまうべく、さまざまな言語で書かれたブログを互いに翻訳しあうためのプラットフォームという着想を得た。

「自分のブログ、もっといろんな国の人に読んでもらいたい」

「最近見つけたこの英語ブログ、周りの日本人にもぜひ読んでほしい」

「この記事、日本語で読みたいな」

OpenLingual は、このような願いを最大限叶えるためのオープンな基盤になり、また、海外のブログを翻訳する、翻訳されたものから海外ブログの情報を活用するという文化をインターネットに根付かせる、そのような大きな野望をもったプロジェクトである。

1 <http://technorati.com/weblog/2007/04/328.html>

2 自動翻訳に関しては 4. 期待される効用の項を参照。

2. 提案内容 - オープンな翻訳の場

概要

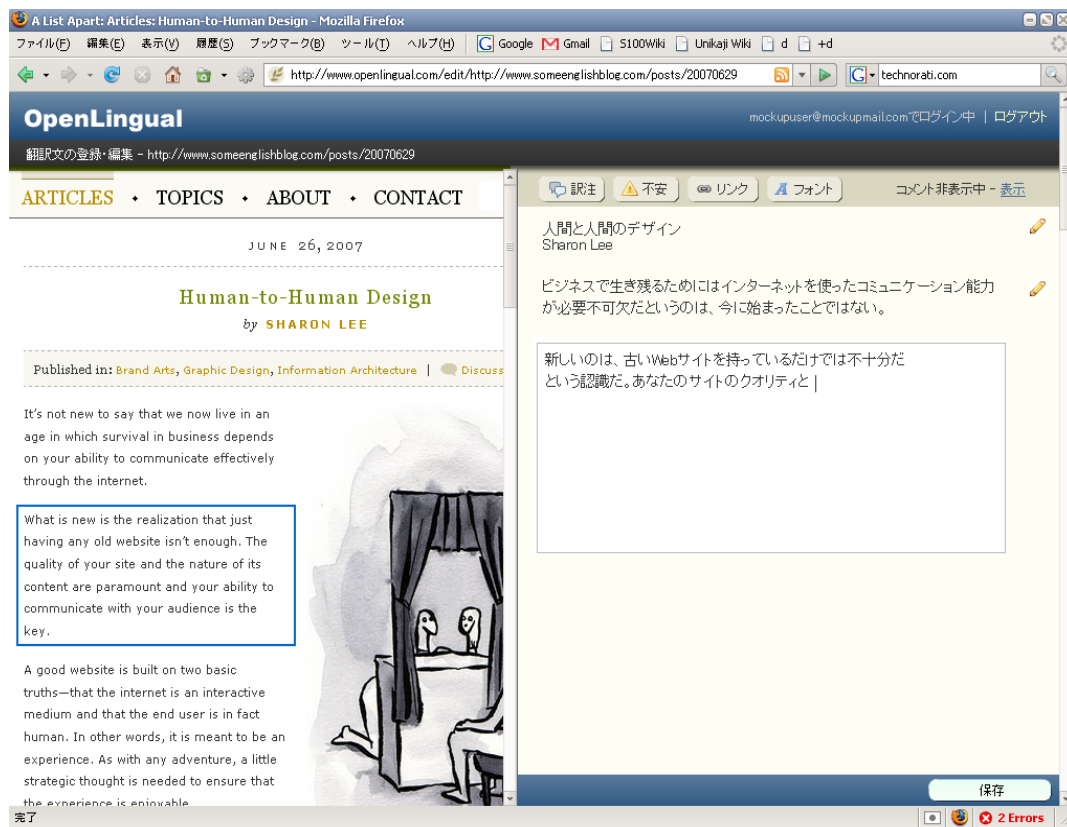
今回提案する「OpenLingual」は、インターネット上のさまざまなページの翻訳文を誰でも作成・掲載できる Web サイトである。また、Web ページの翻訳という行動をより簡単にするための様々な機能を提供し、最高のユーザエクスペリエンスを追求する。

主なターゲットは次の3種類に分かれる。以下ではそれぞれのユーザにフォーカスし、各機能を説明する。

- ・「この記事をもっとわが国の人にも読んでほしい」、すなわち翻訳者
- ・「この記事を母国語で読みたい」、すなわち読者
- ・「この記事をもっといろんな国の人に読んでもらいたい」、すなわち原作者

基本的な機能

任意の Web ページ（具体的にはブログの記事など）に対して任意の言語で翻訳文を作成し、掲載することができる。この翻訳文は誰でも編集可能である。翻訳された文は、後述するブラウザ拡張での同時表示、もしくは OpenLingual のサイト上で閲覧することになる。



翻訳文編集画面イメージ。
左側の原文からセクションを選び、対応する訳文を入力する。

翻訳者を手助けする機能

画面分割インターフェース

画面を半分に分割し、原文を見ながら翻訳文を作ることができる。

インライン辞書検索

メジャーな言語に関してはサードパーティの辞書 Web サービスを利用し、原文をなぞる(ドラッグして選択する)だけでスムーズに辞書検索ができる。

訳注・コメント

任意の箇所に対してコメントを付与することができる。このコメントはページ最後尾などにまとめて付与されるのではなく、あたかも付箋紙を張ったかのように場所を特定して付与することができるため、読者にとっては閲覧時の負担が軽減される。訳者本人がコメントをつけた場合は訳注としての役割を果たすことになるが、コメントを付与するのは訳者本人とは限らず、読者や他の翻訳者がコメントを付与したり返信したりすることも可能である。

不安マーク

よく分からない表現やあまり自信のない箇所に関しては「不安マーク」をつけ、他の翻訳者に助けを仰ぐことができる。また、表現を鵜呑みにしてしまわないようにとの読者への警告になる。翻訳者が必ずしも翻訳のプロであるとは限らない、むしろ素人のほうが多いことを考慮し、敷居を低くするための機能である。これも前述の訳注と同じく、箇所を特定して付与することができる。

不安解決マークおよびデータベース

不安マークの箇所が誰かの助けにより解決された場合には、そのことを示すマークを付与し、同時にログとして蓄積する。これにより、よく登場するが訳しにくい表現などをデータベース化することができ、後の利用に役立てられる。

大衆による編集

前述したとおり、作成された訳文は誰でも自由に編集することが可能である。編集ログおよび差分は全て保存され、閲覧や差し戻しが可能である。

読者を手助けする機能

Web ブラウザ拡張による「翻訳あり」表示

外国語のブログを訪れた際、そのページの翻訳が OpenLingual 上にあることを示す通知アイコンをブラウザ拡張機能を用いて表示する。そのアイコンをクリックすることで、次に述べる対訳表示または OpenLingual 上での訳文閲覧へ画面が切り替わる。

Web ブラウザ拡張による対訳表示

画面を二分割し、左側に原文、右側に翻訳文を表示する対訳ビューを提供する。これは不安マークのついている箇所を即座に原文で確認することができるほか、原文に添付された画像や埋め込まれた動画などのマルチメディアを同時に閲覧することができるという点で有益である。また、デザインやスキンに凝ったブログが多いが、そういった見た目の雰囲気尊重しながら訳文を読むことができる。

なお、以上二つの機能は Firefox など拡張が可能な Web ブラウザか、User Javascript エンジン³の組み込まれた

3 User Javascript エンジン：閲覧中のページに対してユーザ側で任意のクライアントサイドスクリプトを実行できるようにするためのブラウザプラグイン。画面表示の大幅な変更などが可能になる。Opera には初期状態で組み込まれており、Safari や Internet Explorer に対してはサードパーティのものが無償で公開されている。

Web ブラウザに限られる。

翻訳リクエスト

興味を持ったが全部を読む気にはなれないようなサイトやページに行き当たった場合、そのページの翻訳を OpenLingual へリクエストすることができる。その際には翻訳するページの URL、元の言語、翻訳先の希望言語、ページ内容のジャンルなどを指定する。

集まったリクエストは OpenLingual のサイト上で公開され、有志による翻訳を待つことになる。

RSS 配信

特定のサイトへの訳文が追加された際には RSS で更新を通知する。

原作者を手助けする機能

許諾シール

自分のブログをさまざまな言語に翻訳して公開してほしい、もしくは公開されてもかまわないという意思表示として、サイト上に OpenLingual が提供する「許諾シール」を貼ることができる。

これはいわゆる「ブログパーツ」であり、見た目の上では単なる画像だが、実際には Javascript やサーバ側スクリプトを利用して次のような機能が提供可能である。

- ユーザは許諾シールをクリックすることで翻訳文がある場合はそれを閲覧、ない場合は翻訳文を作成またはリクエストできる。
- OpenLingual はシールが貼られた URL をデータベースで管理し、後述する権利の扱いに役立てる。



許諾シールのイメージ

コミュニティとしての機能

プロフィール

ニックネーム、母国語、操れる外国語、語学資格、興味のある分野、自分のブログ URL などをプロフィールとして登録できる。OpenLingual を訪れた翻訳者に対してプロフィールに合致した翻訳リクエストを表示することで、翻訳者にとっては自分の興味ある分野の新たな情報に出会える可能性が増し、読者にとっては翻訳リクエストが実際に処理される確率が上がることになる。また、次に述べるカルマ制を実施するうえでもユーザを識別するプロフィールは必須である。

カルマ

カルマは業（ごう）、すなわちポイント制度である。翻訳者として様々なページを手がけ Blogosphere に貢献したユーザは、その活発度に応じてポイントが与えられる。

カルマを導入することによって次のような効果が期待できる。

- Web リソースを翻訳することに誇りや意義を感じてもらい、OpenLingual、ひいては Web の人力翻訳という行動に愛着を持ってもらう。
- 翻訳文の信頼度の指標になる。すなわちカルマの高いユーザによる仕事はおそらく信頼できるものである。

権利の扱い

著作物の翻訳およびその公開に関する権利の扱いについては各国の著作権法によって規定されているが、基本的に、著作者は著作物に対して翻訳権を所持する⁴。つまり、著作者の許諾があれば著作物を翻訳し一般に公開してもかまわない。

4 日本では著作権法第二十七条。

そこで、OpenLingual では前述した許諾シールによって原作者の許諾を管理する。どの URL に対して許諾があるかはデータベースで管理可能であるため、ページの翻訳を始める際にその URL に対応する許諾の有無を確認したうえで、許諾がある場合のみ翻訳文を公開できるという寸法だ。

つまり翻訳者にとって、翻訳を公開したい場合は、ブログの持ち主に「翻訳していいならシールを貼ってください」と平易に依頼するだけでよい。

そして原作者にとっては、自分のブログにシールを貼るだけで許諾が出せることになる。ユーザ登録などは一切必要ない。

3. プレゼンテーション

OpenLingual は、最近強まっているオープンなコラボレーションの流れに乗る Web サイトとして公開する。立ち位置としては Wikipedia のようなものを想像していただきたい。多くのユーザによる開かれたコラボレーションの場として提供することが狙いであるため、営利目的とせず、収入は必要最低限の運営費用を回収するための広告収入にとどめる。また、時機を見て開発をオープンソース化したいと考えている。

5 公開することだけが問題であって訳文の作成自体には制約がないため、許諾のない時点でも翻訳作業を始めることは可能。

4. 本プロジェクトの意義

発想の原点

私は大学で情報システムを専攻しているが、趣味としてもプログラミング、PC、Webなどの話題が好きで、最新のWeb技術などに関する様々なブログを日常的に読んでいる。海外生活の経験があるので英語で書かれたブログもよく読むのだが、やはりシリコンバレー周辺から発信される技術ネタはとて品質が高いということを感じている。日本のブログもそう負けてはいないが、やはり本場の情報は迫力があるし、なにより日本人にはない発想で物事を捉えていることが多く、とても興味深い。

私の身の回りにはそのような話題を好む友人が自然と集まっているので、私としては是非そのような質の高い記事を彼らに読んでもらいたい。そしてそれについて熱い議論を交わしたい。だが、彼らにとってそれらのブログは敷居が高く、かなり本腰を入れて取り掛からなければ難しいのだ。せっかく有益な情報があるのに読めないなんて、なんともったいないことだろうか。

そこでいくつかの質の高い記事を翻訳して自分のブログで公開しはじめたのだが、私という個人のサイトでその訳文を公開することに違和感を感じた。これはもっとパブリックな場に置くべきではないだろうか？

また、意識し始めると、技術系の英語ブログを翻訳している日本語ブログがいくつか存在することに気がついた。彼らとコラボレーションができれば、すばらしい効果が生まれるのではないかと？

それがWebの共同人力翻訳プラットフォーム「OpenLingual」着想のきっかけである。

なぜ、人力翻訳なのか

現在、自動翻訳ソフトウェアの開発が活発に行われ、Webページの翻訳もワンクリックで行えるまでになっている。そこであえて人力翻訳プラットフォームを登場させる意義とは何か。

それは単純に、自動翻訳は精度が低いということに尽きる。

Webページの自動翻訳は、現状では上手く翻訳できているとは言いがたい。これはために日本語の文章を英語に翻訳し、出てきた英語を再び日本語に戻してみればよくわかる。もはや原形をとどめていないことが殆どである。自動翻訳はどうしても必要な場合に使えば何とか意味は読み取れる、というレベルであり、表現力や読みやすさでは到底人間にかなわない。

もちろん自動翻訳の技術は日々向上していて、人間とさほど変わらない腕前で翻訳ができるようになる日もそう遠くはないのかもしれない。しかし、平たく言えば、それまで待てないのである。数千万ものブログがまるでダムのような勢いで情報を配信し続ける今、最大多数の人が世界中の質の高い情報にありつける国際化プラットフォームを、少しでも早く用意することが必要不可欠だ。

なぜ、共同のプラットフォームなのか

ではなぜ個々のブログではいけないのか。共同のプラットフォームとすることの意義とは何か。

その大きなポイントは、共同作業のしやすさだ。

海外ブログを翻訳したものを掲載している日本語ブログがあると上で述べたが、そのような訳文を見ていると慣用句の意味を取り違えていたり、分からなかった部分を放置していたりということが少なからず見受けられる。私は間違いを見つけた場合すぐにコメントで指摘するようにしているのだが、全ての読者がそうとは限らないし、むしろ少数派だろう。言い方によっては相手を傷つけてしまうのではとの思いから指摘を控える人も少なくないはずだ。

予算内訳

品目	金額	備考
人件費（提案者）	1,200,000	6時間×4日×25週
人件費（共同開発者）	600,000	4時間×3日×25週
Webホスティング費用	18,000	月額約3,000×6ヶ月
備品等の購入	300,000	ノートPC×2台
プロジェクト管理組織費用	500,000	
計	2,618,000	

6. 過去の成果物

Webアプリケーションの開発を行ったことがあり、実際に現在稼働中である。これは新入生に対して同志社大学のサークルを紹介するWebサイトで、サークル員はユーザー登録して自らのサークルを掲載することができる。稼働環境はLinux/Apache/MySQL/PHPで、PHPについてはEthnaというフレームワークを使用している。2007年初頭より開発を開始し、2007年3月から本稼働中である。

「サーカジ」<http://cir.uni-kaji.com/>

7. 提案者について+将来のソフトウェア技術について

私は、美しいものを観たり聴いたり造り出したりすることにこの上ない喜びを感じる。それはたとえば音楽であり、絵画であり、文章であり、メディアデザインであり、商業デザインであり、はたまたソフトウェアデザインであったりする。

媒質はなんでもよい。何かしら人の造るモノには美しさの宿る余地がある。そして、美しさはときとして素敵な仲間を連れてくるのだ。絶妙なシンプルさ、エンターテイメント、セクシーさ、深遠な哲学的思想。これらは美しさと互いに高めあい、モノの放つ輝きをいっそう増してくれる。

美しいソフトウェア？セクシーなソフトウェア？そんなものは果たして存在するのか、と問いたくなるが、逆に考えれば美しくないソフトウェア、セクシーでないソフトウェアなどは身の回りにあふれかえっている。ならばきっとあるに違いない、なければ造れるに違いない、と私は考える。

ソフトウェアの美しさとはすなわち極限まで無駄を省いた仕様、分離性、再利用性、マニュアル不要の直感的なインターフェースだ。ソフトウェアのセクシーさとはすなわちスムーズな動き、人が思わずにやりとってしまう小さな喜びを仕込むことだ。私にとって良いソフトウェアとは、そんなうれしい性質をそろえ持ったソフトウェアを言う。ある機能を果たすだけでは不十分で、それをいかに上手くこなすか、そこにこそソフトウェア開発の意義を感じるのである。

もちろん、演出すべき良さこれだけではないし、それを演出する方法もまだまだあるだろう。これからの自分にとって、新しいフロンティアを開拓するのももちろんのことながら、今ようやく達成しているような機能をさらに美しく、さらに使いやすく、さらにセクシーに実現する方法を模索することが大きな使命であるように感じられる。これはまた、これからのソフトウェア技術一般について期待することでもある。