

辞書作成のリエンジニアリング

岡地 優司・山崎 啓

国立電気通信大学大学院 情報システム学研究科
国立電気通信大学 電気通信学部 電子物性工学科

本論文では、限られた短いスケジュールで辞書を完成する1つの方法が記述されている。大部分の辞書は、多数の専門家に執筆を依頼し、膨大な時間をかけ、それに加えて大人数で多数回の編集会議を経た結果として作成されている。例えば「化学大辞典」(共立出版)は完成まで10年を要し、約百人の編集委員のもと、延べ400人強の専門家が執筆にあたっていた。これでは辞典が完成する頃には「古語辞典」になってしまふことが度々起つる。そこで我々は約2200語からなる情報分野の辞典を、限られた数の専門家達と新進気鋭の大学院生達の力を借りて、期間3ヶ月で完成させることに成功した。

RE-ENGINEERING OF PRODUCTION A DICTIONARY OF INFORMATION

Y u j i O k a c h i , A k i r a Y a m a s a k i

The University of Electro-Communications
Graduate School for Information System
Department of Applied Physics and Chemistry

This paper describes a way to make a dictionary of information science in very tight schedule. Many dictionaries are usually the production of numerous famous professionals under the busy and tedious editorial committees. For example, 「ENCYCLOPEDIA CHIMICA (Publisher : Kyoritsu-Shuppan)」 needed almost 10 years and 400 professionals. Therefore, it needs many years until publication, and the contents and explanations are occasionally inclined to become old-fashioned and meaningless. We have succeeded in preparation of the complete manuscripts to overcome these problems for a novel "Dictionary for Information Science" in only almost three months. Although this dictionary is pocket-book size, it contains above 2200 terms, whose explanations have been written by limited number of special writers including fresh graduate students of proper research area aiming for the beginner's guide and readers' good understanding.

1. 序論

ヒトの集合体において、それぞれに文化と習慣があるように、企業や部署によつてはそれが勝手に言葉を創ったり、方言を使うことが多い。例えば、「カード」という言葉はかつては「パンチカード」を意味していたが、様々なデータベースが注目をあびる現代では「カード型データベース」を指すのが通常である。こうした問題は、企業や部署による多様な言葉が現れる。しかし、企業はその業界でしか通用しない言葉を用いて、業界で一般的な問題であり、そのためには企業に沿うべき言葉を用いる。そのためには、企業は日々の業務で多くの言葉を用いており、それが企業の生産性を高める重要な要素となる。しかし、言葉の変化は常に起こり、そのためには定期的に言葉を更新する必要がある。

そこで、今回は6月20日に講談社より発売予定の「情報マルチメディア辞典」の作成を機会に、実際に辞書を作成するにあたって能率良く作成したか、また、どれほどの効果があったのかを報告する。

辞書は本来、その使命（わからない単語をわかるようにさせるということ）から、単語の意味が大きく変化したり、単語自体が死語になる前に完成させなくてはならない。しかし、現状では辞書の作成に数年を要しているものでも、「短期間で作成できた」といわれている。その結果、辞書が完成する頃には製品自体が一世代前であったり、極端な場合、担当者の顔が全て変わっていたりすることもある。

2. 辞書の存在意義

ところで、辞書を作成することの重要性は以下の点にある。

①テクニカルタームの問題は、その世界について全く知らない人々にとって大きな障壁となることが多く、一部のマニアのみの理解事項として一人歩きし、世間にはなかなか受け入れられないという傾向を生み出す。

②それぞれの業界の新人にとって大きな問題である。新人を迅速に育成して稼ぎ手になってほしい企業にしてみれば、どうしても見逃せない問題である。（この言葉に関する問題はこれから話題になるISO9000にも関係している。）

こうした理由から辞書は本来、その使命（わからない言葉をわかるようにさせるということ）からすると、単語の意味が大きく変化したり、単語自体が死語になる前に完成させなくてはならない。すなわち辞書の使命をまとうするためには

- ①単語の意味自体が変化する前に作成を完了させる必要がある
- ②読んだ人なら誰でも理解できるものである
- ③改訂が可能であること

以上の3点が守られなくてはならないことになる。

3. 今回の辞書作成の背景

今回「情報マルチメディア辞典」を作成することになった背景として以下のことがあげられる。

- 1 中学校、高等学校における新科目「情報基礎」で使えるような、手軽でやさしい辞典がなかったこと
- 2 年々急増する情報処理技術者のニーズ
- 3 海外で標準となってきたISO9000やJISへの対応

そこで、今回は以下のような手順（出版社との交渉などは除く。）に従って、辞書の作成を行った。

- ①辞書に掲載する単語の選定
- ②メンバーの選定
- ③中間段階のサンプル作成
- ④サンプルの評価と修正
- ⑤項目の見直し（見直しと削除）
- ⑥本格的な執筆原稿の作成

4. 今回の辞書作成の手順

①辞書に掲載する単語の選定

辞書で必要な単語のリストは、個人の主觀にとらわれないようにするために、他の本や辞書（複数）を参考として作成する。まず、いくつかの重要な専門書や白書などの巻末についている索引を、表計算ソフトを使ってデータベース化するところから始める。表計算ソフトを利用するには以下の理由による。

- ①ソートが行いやすい
- ②カード型データベースはカードの定義が煩わしく、変更の度にフォーマットの修正を行わなくてはならないが、これはきわめて面倒である。
- ③数千からの項目を扱うため、カード型のデータベースではメモリの大きさに不安を感じる。
- ④簡単に全体を見回すことができる。（マネジメントを容易にする。）

候補となる単語の選定に関しては、編集委員の偏見や好き嫌いを極力排除するため、他の辞書や参考書に力を借りる。今回の場合、内外の約15冊の辞書と数冊の専門書（マルチメディア関係、ワープロ、表計算に関する本：これらは最近始まった中学や高校の情報基礎で登場するソフトに関連する。）で掲載されている単語をそれぞれデータベースに登録した。ここでは、それぞれに関連する分野のあることも考えて、少し多めに採録することとした。（普通の単語でも、分野によっては特定の意味を持たせて使うことが少くないからである。）

次にこのようにして集めた単語をさらに一つのファイルに集めてマージし、単語の頻度を調べる。頻度の高いものは重要度が高いと考えられるので極力残し、頻度の低いものは削除を検討する。

本来、このような辞書の作成にあたっては、編集委員の過去の経験や授業で使った本などを参考に作るため、どうしても時間が非常にかかる上に、単語もかなり偏ったものになりやすい。ここで、他の本をいろいろとマーケティングすることで、辞書としての特徴が多少は失われるかもしれないが、その替りに、選定された単語の中立性と最新の必要語彙の収録という重要な条件は少なくとも守ることが可能となる。

この重複を落とした結果、約2600語以上の単語が集まり、さらに検討の結果、この中から約2400語の単語を執筆対象項目として登録することとなった。

最後に別項参照の単語をその中から探す。一般的に常識でわかる単語は、見つかり次第組み合わせていく。参照項目（「***を見よ」となる語）と被参照の項目（前述の「***」にあたる語）は、自分だったらどちらの項目で引くことが多いかで判断した。この作業により約300語が消えた。同時に、登録された項目の英語名を別枠で登録する。これは後に英字名でソートしたときに類似

語は同じ英字名かかなり似たような英字名になることで判断できると考えたからである。例えば以下のようなケースもある。

エレメント (→素子)	element
素子	element
J C L	job control language
ジョブ制御 (→ J C L)	job control language

②メンバーの選定

多くの辞典類は数多くの執筆者数を誇るのが常のようで、数十人から数百人のリストが巻頭に載っていることも珍しくない。だがこれは、内容のレベルの不均等と原稿の集まりの悪さを当然のこととし、日進月歩の世界においては好適とは考えられない。今回の辞典のような規模と対象を考えると、どうしても少數精銳の執筆スタッフと、かなり強力な編集委員を揃える必要が生じてくる。

そこで、マルチメディアに精通している人、ハードウエア、ネットワーク・オタクなど、メンバーは辞書の幅の広さから多種多様な人物を集めることにした。さらに辞書は、初学者にも分かりやすいことが求められるところを考慮して、カルトの上を行くような知識のありすぎる超大権威の先生は避け、新進気鋭の学生を登用した。この結果、以下のようなメンバーになった。(単なる協力者は除いてある)

- | | |
|---------------------------------|-----------------|
| 1 コンピュータ一般
電通大 電気通信学部 電子情報学科 | 学部生 3名
教官 2名 |
| 2 ハードウエア
電通大 情報システム学研究科 | 院生 1名 |
| 3 マルチメディア、法律関係
新潟大 法学研究科 | 院生 2名 |
| 4 表計算ソフト、ワープロ関係
電通大 | 教官 1名 |
| 5 ネットワーク一般
電通大 情報システム学研究科 | 院生 1名 |
| (合計) 10名 | |

③中間段階のサンプル作成

続いて内容の作成にとりかかる。前述①で作成した項目は、全てが②の分類で分担を割り当てることができるわけでもなく、また分類できたとしても確実にこなしてくれるかどうかも非常に怪しい。そこで項目の分類は特に行わず、執筆担当の項目の選択はもっぱら本人の意志に任せた。(各自の目標数は設定した。)

最初のうちは、元原稿を編集委員サイドでこまめに添削し、それを作った人と検証しながら、それぞれの見出し項目に対する原稿を執筆してもらった。(時には執筆者の誤解などもあり、本人の合意なしに編集サイドで大幅に修正したものもあるが、その分は編集委員が全責任を持つこととして了解を得ている。)

このようにしてそれぞれに分担を決めて執筆を始め、最初の原稿が集まつくると、執筆原稿の内容や説明の詳しさ、何をどこまで書くべきかについて、改めて指針をつくることが必要となってきた。それは以下に示す各事項である。

- 1 概念的、抽象的な単語(例:「互換性」「インターラクティブ」など)は

- その使用文例と実際に文の意味をできるだけ書きそえる。
- 2 略語の場合には何の略であるか、説明をつける。(日本語で)
- 3 類義語が存在するときは、その単語との違いを説明してみる。
- 4 その項目がなんらかの事象を示すものである場合には、事象の内容を説明するだけでなく、事象の起きた理由も考えて書き込んでみる。例えば「誤差」とか、「故障」、「実行」などが該当する。
- 5 製品に関連する言葉や流行語の場合には、その発生に関する説明をつける。

以上のこと留意しつつ、各項目の意味を定義し、それぞれの執筆をはじめることになった。

さらに各自が特定の単語の処理で悩み、その結果生産性が落ちないように項目の予約・予約解除を認めるようにした。

④サンプルの評価と修正

サンプルの評価と修正はメールシステム（E-mail）などを使って、各関係者に意見を求め、これにより執筆原稿中の誤解を防ぎ、誤字脱字を減らすことができた。（遅ればせながら、今回の辞書のモニターを行ってくれる晃華学園高校からも意見をいただいた。）同時に言葉の使い方（用語や言い回し）も統一を図ることにした。以下はその例である。

- 1 同じ意味に取れる「コンピューター」、「コンピュータ」と「電子計算機」と「計算機」では単語の説明中にどの単語を使うべきか?
 → 「コンピューターマッピング」という単語はあっても「電子計算機マッピング」「計算機マッピング」という単語は使われない。そこで「コンピューター」を優先的に使うことにした。「コンピューター」と「コンピュータ」では、後々の修正を考えて前者から後者への修正が容易であることをつけ、前者を標準とした。
- 2 漢数字とアルファベットの数字はいかに区別して使うべきか?
 → 「RS-232C」のようなすでに決められているものは除き、出版社の編集担当者に任せることにした。編集のプロこそが一番よくわかっているに違いないと判断したためである。また、出版社それぞれがシリーズでの統一をとっていることも少なくない。

⑤単語の見直し（見直しと削除）

仕事を開始してから1ヶ月半後、かなりの項目についての原稿が集まってきた一方で、誰も手をつけようとしない項目が残った。さらに、今回の辞書の対象が中学高校生レベルとしてあることを考えると、現在の辞書に掲載しようとしている単語のレベルが果たして本当に合っていると言えるものなのかという疑問の声があがってきた。
 そこで、辞書の各項目の執筆が始まってから約2ヶ月後（この時点で6割以上がほぼ完成の段階に至っていた。）に辞書の項目の見直しを始めた。その結果、約200語が不要であるとして削除された。

今回で対処できなかった単語については、クリアする方法がいくつか考えられたが、結局、今回の辞書の見出し項目からはとりあえず削除し、今後の改訂版の刊行の際の検討課題として残すこととした。

⑥各項目に対する辞書原稿の作成

あとは、全体の数字を見ながら、着々と進めることになる。中間集計をとりながら仕事の進み具合を見ることになるが、遠隔地とのやりとりなどの処理もあって、リアルタイムの集計はかなり難しく、出てくる結果はかなり前のデー

タになりがちであった。

それでも結局、ほぼ予定通りに仕事は進み、9月最後の週に講談社の編集担当者にソート済み、英文索引を付与した完成原稿を渡すことができた。

5. 最後に

今回の辞書の作成を通じて、教訓となつたことは

- ①就職戦線が長引きすぎて、人のアサインをしてもその通りに作業が進まなかつたため、計画が大幅に乱れた。
- ②短期間であったため、若手の執筆者の教育に時間が割けず、その結果、文体や語句の統一、禁句の訂正などによる編集委員の負担が増えた。
- ③電通大の学生の文章を書く能力が著しく劣り、スペースの速い人と遅い人の生産性の違いは15倍以上異なる事実を計算に入れていたこと。

以上の3つが考えられ、また、今回の辞書が最終的にはどれほどの出版部数出たのかはまだ不明であるが、少なくとも今回の辞書作成を通じて、短期間に内容のある辞書を作成する秘訣をまとめることができた。その概略は以下のようなものである。

- ① 編集委員の中に、実際に辞書を執筆した経験のある人がいること。
- ② 担当する項目は義務でなく権利から生じるものであるから、できないと判断したら速やかに放棄（辞退）して、他の人に任せられるようなシステムをつくること。
- ③ 編集委員が絶対的な権力を握る。（単語の割り振りや執筆原稿の訂正是、本人から許可を得なくとも実施できる。「大権威の○○氏の書いたのだから、直すことなどおそれ多くて」などというのでは、絶対にいい辞典は出来ない。）
- ④ 執筆メンバーの決定は編集委員の独断と専権で行える。
- ⑤ 表計算ソフトを使っての項目の整理。
- ⑥ 辞書利用者の対象がある程度明確に把握できていること。
- ⑦ マネージャーの役目は第一線の人がやりやすい環境を作ることと肝に命じること。（会議が頻繁にひらかれてばかりいるのでは、本職をやっている時間がなくなる。）

6. 実績比較

実績比較

	情報マルチメディア辞典	化学大辞典 (共立出版)
編集委員	4人	100人
執筆者 (協力者除く)	6人	460人
項目数	2200項目	70000項目
一人当たり	220項目	130項目
作成日数	1/4年	10年
(生産性	65	: 1)