

ソフトウェア開発における用語の木の作成について

後田尚輝 潘建華 伊藤正樹 吉川直邦! 斉藤直希、小川清!!
!大同工業大学 !!名古屋市工業研究所

はじめに

ソフトウェア開発において、開発環境、開発対象の違いにより、用語の意味が異なる場合がある。ソフトウェア開発の効率を妨げる要因の一つである。コンピュータ関連の雑誌・書籍では、略語、カタカナ語が多用されている。固有名詞と標準用語とが混在し、用語の示す範囲が、文脈によって異なることがある。略語には同一の表記で、複数のフルスペルに対応しているものがある。類似の表記で、同一の分野で異なる意味の略号もある。カタカナ語では、文脈に応じた使い分けがされている場合には、厳密な意味を持つ場合と、意味の重複した用語が存在する場合もある。これらの用語のソフトウェア開発するにあたって用語の辞典を作る必要がある、用語の意味はもちろんのこと、それに加えそれに関連する単語も調べていかないといけない。

1 用語の木の作成の方法

1.1 英語文献からの作成

1) 単語切り出し文献の特定

標準文書、仕様定義文書から単語を切り出す。用語体系が違う文書の場合には、独立に切り出し、それぞれを比較するようにする。

2) 栄単語を抽出

英文の場合には、全単語の集計表を作成する。大文字小文字を変換し、記号を削除し、AWK のスクリプトで単語を切りだす。

3) 略号のフルスペル付与

略号は、そのフルスペルが何かを調べる。英日の機械翻訳ソフトにかけ、単語の切り出し方、用語の文脈依存性を検討する。文脈によっては略号として用いているものと、フルスペルとして用いているもの、プログラム上の定数、変数、関数として用いているものがある。

Making Keyword Tree on software development
Ushiroda Naoki, Han Kenka, Ito Masaki,
Yoshikawa Naokuni!,
Saito Naoki, Ogawa Kiyoshi!!
Daido Institute of Technology
Nagoya Municipal Industrial Research
Institute

例)

DSR (Device Service Routine) = DIC 内のサービスコールを実現するためのルーチン[5]

DSR (Distributed Speech Recognition) = 携帯電話や PDA などの携帯機器から、分散処理技術を使って音声で直接アプリケーションを操作できる技術の名称

4) 専門用語かどうかを判定する。

Stop word だけではなく、一般用語と考えられるものは、用語辞書から除外するため、専門用語かどうかの判定をする。

5) 固有名詞、規格用語かどうかの判定

登録商標をはじめとする固有名詞、定義のある規格用語かどうかを判定し、引用を付与する。単語によっては、規格用語と一般用語が混在している場合もある。

6) 用語の包含関係を調べる。

包含関係には、

無関係 A and B =

上下関係 A and B = A

並列関係 A

抽出してきた内容にあった意味を調べる

類語・関連語を調べる

わかりにくい用語は図にまとめてみる

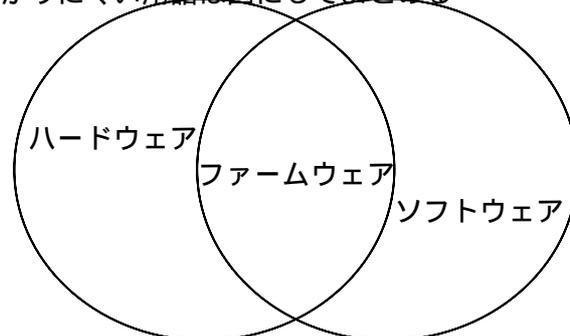


図1、ハード、ファーム、ソフトの関係

1.2 日本語文献からの作成

日本語の場合には、単語の切りだしが難しい。電子ファイルであっても、簡単な方法が利用しづらい。テキスト処理として、漢字、ひらがな、カタカナ、アルファベットの区切りで切り出して一次資料とする。

2 編集方法

2つの概念の共通部分をどのように編集するとよいかを検討した。以下、図1の概念に基づいて具体例で説明する。

2.1 関連語として定義

何か関係のある語は、関連後 (RT: Related Term) として分類しておけばよいので、用語の意味がよくわからない段階でも、分類可能である。RT が膨大になる可能性があるため、RT が何語以上になったら、再分類するという規則を作っておくとよい。

ハードウェア

RT ファームウェア、ソフトウェア
ファームウェア

RT ソフトウェア、ハードウェア
ハードウェア

RT ソフトウェア、ファームウェア

2.2 両方の下位語として定義

上下関係が明確であることが理解できれば、下位語 (NT: Narrower Term) として整理する。上位語 (BT: Broader Term) を記載するかどうかは、全体の用語の量による。

ハードウェア

NT ファームウェア
ソフトウェア

NT ファームウェア
ファームウェア

BT ハードウェア
BT ソフトウェア

2.3 AND による表記

各概念に、共通部分との AND を表記して、掲載する。利点としては、上位概念のどのような特性を持った部分集合であるかがわかる点である。

ハードウェア

AND ソフトウェア ファームウェア
ソフトウェア

AND ハードウェア ファームウェア

電子回路のハードウェアを実現するソフトウェアには、CAD がある。回路の作り方のソフトウェアはハードウェアのためのソフトウェアであって、ハードウェア AND ソフトウェアという定義の仕方も可能である。

2.4 is-a, has-a 関係による定義

2つの概念関係のうち、is-a の関係と has-a の関係を明示する。逆の関係は a-si, a-sah で表記する。

ハードウェア

a-si ファームウェア
ソフトウェア

a-sah ファームウェア
ファームウェア

has-a ソフトウェア
is-a ハードウェア

ファームウェアは、ソフトウェアを含んではいるが、実体としてはハードウェアの一部であることを表現できる。

3 まとめと今後の課題

商品名、サービス名によっては内容を定義していないものがある。定義の変更を明確にせず、内容を変えている場合がある。そのため理解するのが困難な場合がある。例えば、OS とデバイスドライバの関係は、両方ともソフトウェアの下位語として定義可能である。パッケージとして販売されている OS には、デバイスドライバからいくつかのアプリケーションソフトウェアまで同梱されていることがある。そのため、購入者にとっての OS とは、OS の箱に入ったものすべてである。アプリケーションソフトウェア、OS、デバイスドライバのすべてが同梱されているハードウェアの購入者には、これらの区分には意味がないかもしれない。

現在、μITRON にでてくる単語を抽出し用語辞典を作成し、組み込みソフトウェア開発における用語を分析・分類し、ソフトウェア開発シソーラスを作成しながら、システム的设计を行っている。特に、OS とデバイスドライバ、アプリケーションソフトウェアの境界とその設計方法、実現方法について、Toppers プロジェクトの成果と比較して検討中であり、システムごとに世界の異なる組み込み系の開発、教育に役立てることが課題である。

「参考文献」

- [1] データベース検索技術者試験・傾向と対策, 小川清・安藤彰敏編著, 紀伊国屋書店
- [2] 言語教育モデルによるプログラミング教育ツールの設計, 青木賢太郎, 小川清, 廣瀬哲也, 言語処理学会年次大会, 1997
- [3] データベース検索技術シソーラスの作成, 小川清, 山崎貴和, ドクメンテーションシンポジウム, 1995
- [4] 組み込みソフトウェア技術者教育の課題と方向性, 小川清, 渡部謹二, 斉藤直希, 電気関係学会東海支部, 2002
- [5] μITRON4.0 仕様, ITRON プロジェクト