第 6 章

基本システム設計と実装

6.1 設計方針

意味の数学モデルを実現するにあたって，意味の数学モデルそのものの実験を行う環境の作成だけでなく，意味の数学モデルを適用可能な，応用的な実験を行う環境の作成を目標とした．そのために，処理の単位をモジュールとして分け，必要なモジュールを組み合わることで，目的に応じた処理を可能とした．

モジュールに関しては，下の 2 階層に分割して作成を行った (図 6.1)．

(1) 実行プログラムの単位である「プログラム・モジュール」

(2) C 言語の関数ライブラリである「関数モジュール」

本実現では，意味の数学モデルのシステムは，複数のプログラム・モジュールを組み合わせて実現され，また，プログラム・モジュールは，複数の関数モジュールを組み合わせて実現されている．

6.2 プログラムモジュール

数千の単語を利用して意味的連想検索を実現した場合，一般的なワークステーション上で計算を行うと，固有値分解などの処理には数時間から数日間必要である．処理の単位を明確に分割し，それぞれの処理結果を保存し，そして，再計算の必要な場合には，前回の処理結果のうち，使用できるデータをなるべくそのまま使用すれば，全体の処理時間を短縮することができる．この作業に “make” を使用すると，データの依存関係の記述から，再計算が必要なデータを作成するモジュールを自動的に選択し，実行を行うので，素直な実装が可能である．

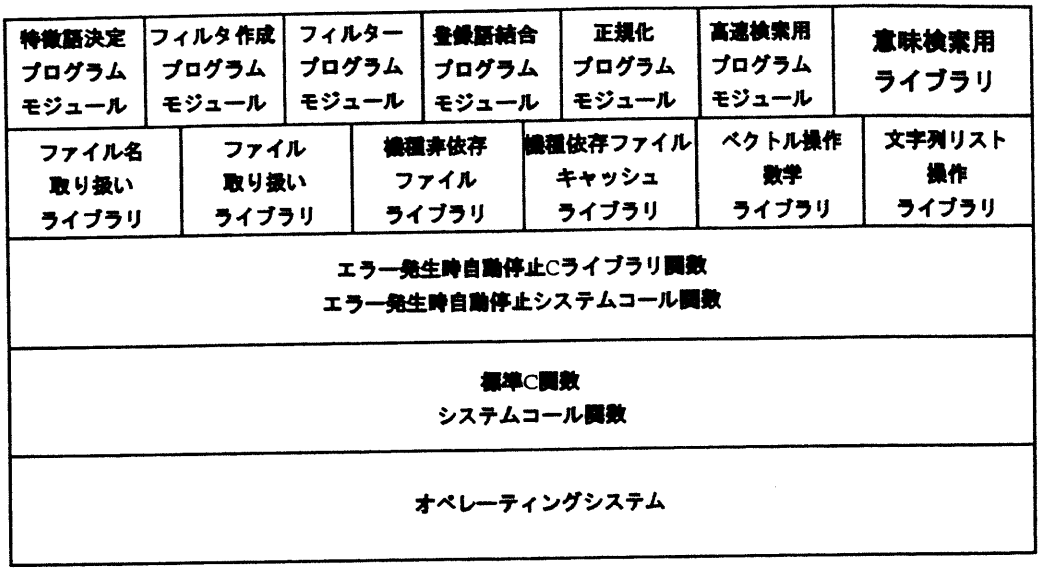

図 6.1: 意味的連想検索システムにおける階層モデル

処理単位でデータを分けることは，対象が異なる処理を行う場合に，前段階の処理結果を共通に利用可能な場合にも有効である．例えば，同一のイメージ空間の上に，異なる連想検索対象を射影する場合には，固有値分解などのイメージ空間の作成を，再度行う必要はない．次に，処理の単位を分割すると，それぞれの処理を得意とする計算機での，分散処理が可能である．

異機種分散環境に対応させるため，バイト・オーダや浮動小数点の，内部表現の相互変換が必要である．そのためには，モジュール間のデータの受渡しには，機種依存の無い標準フォーマットのデータファイルを使用することとした．標準フォーマットには Sun の XDR ライブラリを使用した．XDR ライブラリは，Sun RPC で使用されているフォーマット変換ライブラリで，様々な機種に移植されている．

6.3 全体の機能

意味的連想検索では 3 種類のデータを扱う．

(1) イメージ空間生成のためのデータ

(2) 検索対象語のためのデータ

(3) 検索キーワードのためのデータ

(1) は，イメージ空間生成のためのデータの流れである．英英辞書の定義を利用して特徴を抽出し，A 行列を作成し，固有値分解を行い，イメージ空間を生成する．(2) は，検索対象となるデータの流れである．検索対象語の特徴付けを行い，イメージ空間に射影し，高速な意味的連想検索のアルゴリズムのために，意味素毎のソートを行う．(3) は，キーワード，文脈語に使用される語のデータの流れである．3 つの処理が終了すると，意味的連想検索の準備が整う．ここまでのデータの流れを，図 6.2 に示した．

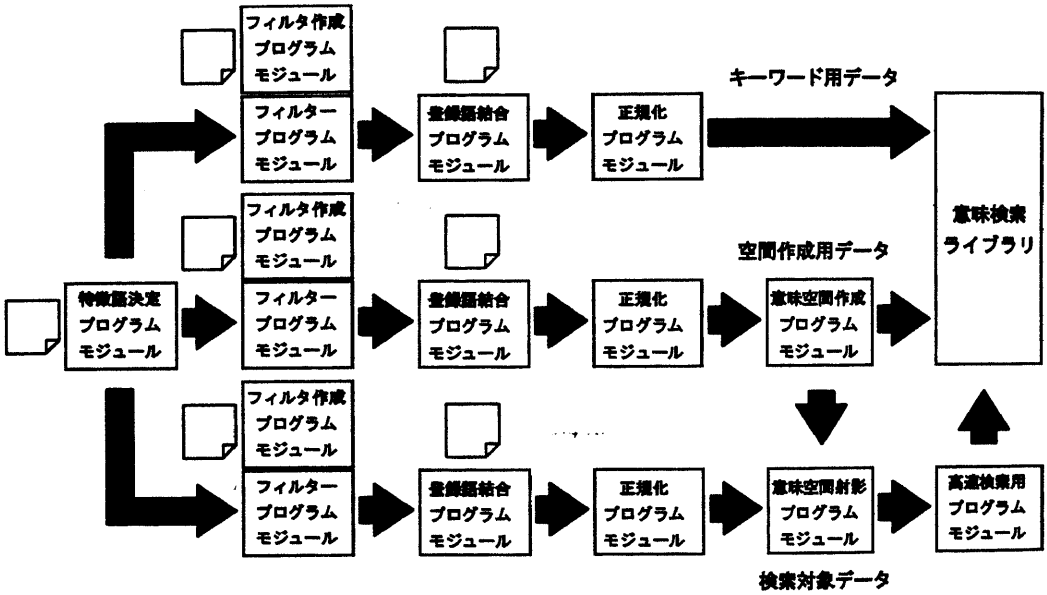

図 6.2: 意味的連想検索におけるデータの流れ

6.4 処理の流れとモジュール

全体を，以下の 6 つのモジュールに分割した．

- フィルター群を通して A 行列を生成するモジュール
- A 行列を加工するモジュール
- A 行列からイメージ空間を生成するモジュール
- A 行列をイメージ空間に射影するモジュール
- 射影された単語群を意味素毎にソートするモジュール
- 意味的連想検索を行うモジュール

6.4.1 フィルター群を通して A 行列を生成するモジュール

入力テキストを，数値行列に変換するモジュールである．基本的な処理は，入力テキストに出現した特徴語に 1 を立て，その他は 0 にすることである．反対の意味で，特徴語が出現した場合には -1 とする．入力テキストに，辞書の定義を利用すると，入力テキスト中の特徴語は，様々な活用形で出現する．活用形を原形に戻すために，入力テキストに対して，多段のフィルターを通すことで対応する．フィルターの動作は，活用形と原形の対応表を参照して，活用形を原形に置き換えることで行われる．

本実装では，フィルター群を通した後は，必ず原形の特徴語になることに注目し，予め，フィルター情報を逆順に参照し，全ての活用形から原形の特徴語へのマッピング情報を作成した．この情報をキャッシュとして保存しておけば，入力テキスト中に出現した単語に対して，マッピング情報を 1 回参照するだけで，多段のフィルターを通すのと同等の変換が可能になる．数千語分の入力テキストは，多くの入力ミスを含んでいたので，フィルターにかけ直す作業はかなりの回数に上った．この様な状況においては，フィルター・キャッシュは非常に有効に働いた．

6.4.2 A 行列を加工するモジュール

例えば，辞書中による単語の定義においては，「何々という単語の何番目の意味が何々である」という記述がなされている．しかし，ユーザーが意味的連想検索を利用する時に，「何々という単語の何番目の意味」という指定の方法は，好ましいとは言えない．意味的連想検索においては，複数の意味が混ざった単語を使用しても，文脈を指定することにより，文脈に応じた解釈が可能である．そこで，ユーザーが指定する単語は，同じ表記を持つ単語の意味を混ぜ，1 つの単語として作成しておく．この様な作業を行うモジュールが，本モジュールである．

6.4.3 A 行列からイメージ空間を生成するモジュール

A 行列の相関行列を作成し，固有値分解を行い，イメージ空間を生成するモジュールである．実装では，A 行列が素行列になる傾向があることを利用した高速化を行っている．他のモジュールと比較すると，このモジュールの処理に最も時間がかかる．

6.4.4 A 行列をイメージ空間に射影するモジュール

生成されたイメージ空間に，単語群を射影するモジュールである．このモジュールの実装にあたっても，A 行列が素行列になる傾向があることを利用した高速化を行っている．

6.4.5 射影後のベクトルをソートするモジュール

高速な意味的連想検索を行う準備として，各意味素毎に，要素の大きさに応じて，検索対象語群のソートを行うモジュールである．

6.4.6 意味的連想検索を行うモジュール

このモジュールは，実際に意味的連想検索を行うモジュールである．意味的連想検索の機能を，一通り行うことが可能である．実際には，後で述べる検索ユーティリティ・ライブラリを呼び出すものとなっている．

6.5 関数モジュール

関数モジュールは，プログラム・モジュールを作成するための，C 言語の関数ライブラリである．関数モジュールには，次のものがある．

- エラー終了付き関数ライブラリ
- 文字列リスト処理関数ライブラリ
- ツールボックス・ライブラリ
- 数学関数ライブラリ
- キャッシュ・ライブラリ
- その他のライブラリ
- 検索ユーティリティ・ライブラリ

6.5.1 エラー終了付き関数ライブラリ

意味的連想検索の本実装では，意味的連想検索の準備が終了するまでは大量のデータ処理に長時間かかるため，バッチ式に処理を進めている．エラーが発生した時は，誤っ

た答を後のモジュールに渡すのを防ぐために，速やかに異常を報告して，終了すべきである．このライブラリは，標準の関数ライブラリを呼びだし，エラー発生時にはエラーを表示して，プログラムを終了させるだけのライブラリである．このライブラリの呼び出しは，通常の関数名の最後に，大文字の E を付けることで行われる．このライブラリを使用することにより，モジュールのプログラムを簡潔に書くことが可能である．

6.5.2 文字列リスト処理関数ライブラリ

プログラム・モジュールの中には，テキスト・ファイルの読み込みを必要とするものがある．本ライブラリは，読み込みファイルの単純な構文解析を行い，その結果をリストにして返すものである．意味的連想検索のプログラム・モジュールの作成にあたっては，ファイル操作や，単純な構文解析などの処理に，時間と手間を取られるべきではない．そこで，この関数ライブラリでは，基本的な機能だけでなく，応用的な，単純な構文解析の機能までサポートしている．文字列の長さと，記憶領域に関わるバグを防ぐため，文字列を扱うファイルは全て，仮想記憶空間にマッピングし，そこへのポインタで文字を扱う．

6.5.3 数学関数ライブラリ

意味的連想検索の実現では，数学の演算を多用する．ベクトルの内積などの，一般的な数学関数を用意したものが，数学関数ライブラリである．

6.5.4 キャッシュライブラリ

異機種分散環境に対応するため，本実装では，モジュール間のデータの受渡しに，標準フォーマットのデータファイルを使用する．しかし，内部では，ローカル・フォーマットでデータを扱わなければ，計算を行うことができない．そこで，標準フォーマットからローカル・フォーマットへの変換を行うのだが，異なるモジュールが，同一のデータを扱う場合，同一のデータに対して，再度変換を行うのは効率が悪い．このライブラリは，データファイルのフォーマット変換と，キャッシングを行うライブラリである．

6.5.5 その他のライブラリ

意味的連想検索の実装に必要なライブラリで，検索ユーティリティ・ライブラリに含まれず，上に挙げたライブラリにも含まれないライブラリは，ここに含まれる．

6.5.6 検索ユーティリティ・ライブラリ

意味的連想検索は，様々な応用分野が考えられる．それらの応用アプリケーションから，意味的連想検索を呼び出す関数群が，このライブラリである．検索を行う意味空間の指定，文脈を表す部分空間の選択，意味空間のしきい値である ε_s の指定，検索キーワードの指定，検索対象語の指定，そして，意味的連想検索を行う関数が含まれている．