

awk から xgawk へ

日本 GNU AWK ユーザー会



awk から xgawk へ

■ はじめに

本日は「日本 GNU AWK ユーザー会」のブースに来ていただきありがとうございます。

今年リリースされたばかりの xgawk 3.1.6 の機能を主に紹介させていただいています。ご不明な点はスタッフまでお気軽にご質問ください。

また、本日の内容は後日以下の URL で公開予定です。

- <http://gauc.no-ip.org/osc2008/>



awk から xgawk へ

■ awk の誕生

ご存知のように awk は 1977 年に Alfred V. Aho, Peter J. Weinberger そして Brian W. Kernighan の 3 人によって作られ、彼らの頭文字を取って awk と名づけられました。

現在、オリジナルの awk が使われることはなく、1985 年に拡張された nawk (one true awk) が主に用いられています。



awk から xgawk へ

■ one true awk から gawk へ

1985 年にオリジナルの awk から関数定義などを加えた one true awk がリリースされ、そのプログラムの簡便さから多くの人に受け入れられることとなります。

一方、翌年の 1986 年には GNU 版である gawk が開発され、オープンソースとしての awk として広まっています。gawk には多くの GNU 拡張がなされ、one true awk との互換性を保ちつつ、使いやすいのが特徴です。



awk から xgawk へ

■ gawk から xgawk へ

近年のインターネットやデータベースの普及に伴い、従来の awk のように 1 行単位で処理できない構造体テキストが増えてきました。XML のような構造体に対しても awk を使いたいという要望は多く、2003 年のクリスマスに xgawk (当時 xmlgawk) が誕生します。



awk から xgawk へ

■ awk の種類

awk にはさまざまな種類があります。ここでは代表的なものを紹介します。

- nawk : Kernighan によってメンテされています。
- mawk : BSD, Debian などでは標準です。
- gawk : 一般的に awk といえば gawk を指します。
- xgawk : gawk の拡張版であり、本日のメインです。
- jawk : Java による awk の実装です。



xgawk 3.16 の紹介



awk から xgawk へ

■ 時間関数拡張

nawk などには時間関数がないため、date コマンドとの組み合わせや srand() の戻り値を使って計算したりする必要がありました。gawk 以降拡張され systime() 関数や strftime() 関数も加わりました。

xgawk では sleep() 関数が加わったことで、さらにアプリケーションの利用の範囲が増えていきます。



awk から xgawk へ

■ XML

xgawk の最初の拡張である XML パースは libexpat を用いています。gawk で使用できる TCP / UDP の通信を使って、Web 上の RSS 情報を取り寄せ処理することも可能になっています。同様に Twitter のクライアントのようなものも作ることができます。

ただし、well formatted である XML に限られていますので、使用する際には注意が必要です。



awk から xgawk へ

■ MPFR

MPFR により任意精度の計算も可能になりました。従来の倍精度浮動小数点演算では困難であった精度の必要とされる計算も可能になっています。



awk から xgawk へ

■ PostgreSQL

xgawk の拡張として libpq を用いて PostgreSQL へアクセスすることができます。

これにより、awk で扱えるデータベースは CSV, TSV のような行単位のテキストだけでなく、PostgreSQL にもアクセスして処理が可能になっています。



awk から xgawk へ

■ GD

GD の拡張機能が加わりました。今までグラフィックスを扱えなかった awk にとって大きな進歩ですが、使いどころが悩ましい拡張です。



awk から xgawk へ

■ xgawk は使われているか？

xgawk はいくつかの Linux ディストリビューションで使われているようです。

また日本では Anarchy Golf で使われていますが、これらの拡張は行なわれていないようです。

OSC2008 Tokyo/Spring の Web ページは最新版 xgawk で動作させています。

• <http://gauc.no-ip.org/osc2008/>



awk から xgawk へ

■ 求められる拡張

xgawk で多くの拡張が行なわれてきましたが、我々日本人にとって重要な拡張がなされていません。それは文字コード変換です。xgawk には XML の場合のみ文字コード変換が行なわれますが、通常処理に反映されていません。これは今後の課題と言えるでしょう。

また、gawk 3.1.5 からバイト単位の処理から文字単位の処理になり、文字のバイト数が取得できないため、日本語で書かれた文書の整形などの処理が難しくなっていました。これも今後の課題といえます。



awk から xgawk へ

■ gawk の日本語化

以前 jgawk, gawk + multi byte extension といった日本語を awk で扱うための拡張がされていた時期がありました。これらは日本人には非常に受け入れやすいものであり、awk ブームの火付け役でもあります。

しかし、オープンソースでは日本独自のパッチを当てるのではなく、upstream に説明して修正してもらうのが本道であると考えています。中には gawk そのものではなく、glibc に依存するようなものもありますが、対応していきたいところです。



awk から xgawk へ

■ gawk の日本語化

以前 jgawk, gawk + multi byte extension といった日本語を awk で扱うための拡張がされていた時期がありました。これらは日本人には非常に受け入れやすいものであり、awk ブームの火付け役でもあります。

しかし、オープンソースでは日本独自のパッチを当てるのではなく、upstream に説明して修正してもらうのが本道であると考えています。中には gawk そのものではなく、glibc に依存するようなものもありますが、対応していきたいところです。



awk のコミュニティ



awk から xgawk へ

■ コミュニティ

日本語を扱う上での問題や日本で使う人の手助けとなるように「日本 GNU AWK ユーザー会」を 2004 年に立ち上げました。未だに組織としては未熟ですが、gawk や xgawk に対して何件かバグ報告や改善要望を出したり、翻訳活動を行なっています。

イベントとして、毎年の LL Conference には協賛または参加しています。

ご興味があれば、ぜひ参加してみてください。



awk から xgawk へ

■ 翻訳についての注意

gawk 3.1.6 から GPL Ver. 3 になりましたが、po ファイルに反映されていないため、日本語ロケールだと GPL Ver. 2 の表示になってしまいます。ご注意ください。

こうした翻訳活動はコミュニティとして進めていきます。



awk から xgawk へ

■ さいごに

これから awk はどうなっていくのでしょうか？ マスコットとして用いられている great auk (オオウミガラス) は地面に落ちている石を拾うよりも簡単に捕獲することができたため絶滅してしまいましたが、awk は gawk / xgawk と進化していきます。

awk は常に、プログラムの良きツールとして、プログラムそのものとして、利用していただけのような awk に進化していくことと願っています。

