

---

# Google Summer of Code 2008 Project Proposal: Porting BSD-licensed Text-Processing Tools from OpenBSD

Gábor Kövesdán <gabor@FreeBSD.org>

Copyright © 2008 Gábor Kövesdán

## Applicant's Information

- Name: Gábor Kövesdán
- Nationality: Hungarian
- Resident of: Budapest, Hungary
- Birth date: 2 Aug, 1987
- Gender: Male
- Mailing address: Pozsonyi u. 2/b, Budapest, 1045
- Phone: +36 30 434 04 06
- E-mail: <gabor@kovesdan.org>
- WWW: <http://www.kovesdan.org>

## Project Description

Proposed mentor: Diomidis Spinellis

FreeBSD currently uses the GNU versions of the command-line text-processing tools: **grep**, **sort**, **diff**, **patch**, **sdiff**. The goal of this project is to port the BSD-licensed versions of these tools from OpenBSD and optimize the performance where possible, provide standard-conformance and handling of wide character sets. The man pages also need to be revised and completed.

## Project Deliverables

This project has already been selected once for a Summer of Code project, which shows its importance. In the end, the student withdrew his application and the project was left undone. It would be very useful to have our own versions of these utilities for many reasons:

- It would make us independent from the external vendor and its development method.
- This independence would also give us more opportunities to customize and optimize the code without the inconvenience of having to handle the merges from the vendor every time a new version comes out.
- The BSD-license suits our development philosophy better.

## A more detailed TODO

I will describe here my tasks, which I will need to complete according to my preliminary investigations.

### Overall items

- Check GNU compatibility with our GNU versions in the base system. This is the most important item, later the actual GNU versions can be checked as a source of extra features.
- Check POSIX-conformance.
- Cleanup code to respect style(9).
- Write shell scripts for the ported utilities to run regression test.
- Find a method to collect which options are used by the Ports Collection so that we can measure the progress of the compatibility.

### grep

| GNU option           | Status in the port           |
|----------------------|------------------------------|
| --label              | UNIMPLEMENTED                |
| --null               | UNIMPLEMENTED                |
| --color / --colour   | UNIMPLEMENTED                |
| -D / --devices       | UNIMPLEMENTED                |
| -H / --with-filename | UNIMPLEMENTED / INCOMPATIBLE |
| -J / --bz2decompress | UNIMPLEMENTED                |
| -d / --directories   | UNIMPLEMENTED                |
| -m / --max-count     | UNIMPLEMENTED                |
| -o / --only-matching | UNIMPLEMENTED / INCOMPATIBLE |

- Teach **grep** to use libarchive to handle .gz and .bz2 files.

### sort

| GNU option                  | Status in the port |
|-----------------------------|--------------------|
| -g / --general-numeric-sort | UNIMPLEMENTED      |
| --help                      | UNIMPLEMENTED      |
| -M / --month-sort           | UNIMPLEMENTED      |
| -S / --buffer-size          | UNIMPLEMENTED      |
| --version                   | UNIMPLEMENTED      |

### diff

| GNU option                 | Status in the port |
|----------------------------|--------------------|
| --ignore-file-name-case    | UNIMPLEMENTED      |
| --no-ignore-file-name-case | UNIMPLEMENTED      |

| GNU option                | Status in the port |
|---------------------------|--------------------|
| --strip-trailing-cr       | UNIMPLEMENTED      |
| --normal                  | UNIMPLEMENTED      |
| --left-column             | UNIMPLEMENTED      |
| --suppress-common-lines   | UNIMPLEMENTED      |
| --GTYPE-group-format      | UNIMPLEMENTED      |
| --line-format             | UNIMPLEMENTED      |
| --LTYPE-line-format       | UNIMPLEMENTED      |
| --tabsize                 | UNIMPLEMENTED      |
| --unidirectional-new-file | UNIMPLEMENTED      |
| --from-file               | UNIMPLEMENTED      |
| --to-file                 | UNIMPLEMENTED      |
| --horizon-lines           | UNIMPLEMENTED      |
| --speed-large-files       | UNIMPLEMENTED      |
| --help                    | UNIMPLEMENTED      |
| --text                    | UNIMPLEMENTED      |
| --ignore-blank-lines      | UNIMPLEMENTED      |
| -c                        | INCOMPATIBLE       |
| --ignore-tab-expansion    | UNIMPLEMENTED      |
| --show-function-line      | UNIMPLEMENTED      |
| -u                        | INCOMPATIBLE       |
| -W/--width                | UNIMPLEMENTED      |
| -y/--side-by-side         | UNIMPLEMENTED      |
| -v/--version              | UNIMPLEMENTED      |

## Timeline and Schedule

I will firstly focus on **grep**, **sort** and **diff** so that they are completed and ready to enter CVS. Only after that I would work on **patch** if I have some remaining time at the end of the program. It seem that there is an ongoing effort to port **sdiff** to FreeBSD. Another extra projects might be checking what other BSD-licensed tools can be ported from OpenBSD or NetBSD (e.g. **bc**) or writing BSD-licensed command line tools from scratch, which are present in other UNIX-like operating systems, but not in FreeBSD. (e.g. **pstree**, **tac**)

- May 26 - Jun 10: Study the code, get a deep understanding of the inner working of these tools
- Jun 10 - Jun 25: Feature completion of **grep**, documentation completion, considering optimization opportunities
- Jun 25: Release **grep** for testing and review
- Jun 26 - Jul 10: Feature completion of **sort**, documentation completion, considering optimization opportunities, process community feedback about **grep**
- Jul 10: Release **sort** for testing and review

- Jul 10 - Jul 25: Feature completion of **diff**, documentation completion, considering optimization opportunities, process community feedback about **sort**
- Jul 25: Release **diff** for testing and review
- Jul 26 - Aug 11: Process community feedback about **diff**, spare time for catch-up in case of a slower progress, work on related extra items in case of a faster progress

## Working Experience and References

I have been a FreeBSD user for more than four years and a committer (doc/ports) for more than one year now, thus I have a wide overview about how the userland tools work under FreeBSD. I have already looked at the text-processing tools from OpenBSD and have modified them slightly so that they build under FreeBSD. I have committed them to the Ports Collection. I have a strong ANSI C knowledge, which I want to widen in order to better cover the programming of the standard POSIX environment. I have already participated successfully in the Summer of Code program twice. For a more detailed description about my experience and knowledge, please refer to my CV: [HTML](#), [PDF](#), [RTF](#), [PS](#), [TXT](#)