

Exhibit 1

EXHIBIT P

DISCLOSURE OF INVENTION

1. Title of Invention: Information Filtering using Machine Learning
(Use technical terms and not proposed trademarks)

2. Inventor(s): Ken Lang

a. print or type name Ken Lang signature
date

department SCS phone 421-3265
% of contribution 100%

b. print or type name signature
date

department phone
% of contribution

c. print or type name signature
date

department phone
% of contribution

d. print or type name signature
date

department phone
% of contribution

3. Brief description of invention. How does this invention relate to new processes, machines, compositions of matter, manufactures, etc.? (please feel free to use additional sheets to elaborate and to attach sketches, drawings, photographs and other materials that help illustrate the description)

The work is the Ph.D. thesis of the inventor, Ken Lang. In this research project, called "NewsWeeder", Ken has applied traditional machine learning techniques to the domain of learning how to filter netnews articles on the internet. A World-Wide Web interface has also been built as a front end to the system.

4. External Sponsor(s) Digital Equipment Corporation Contract or Grant Number(s) 1-41411
(Corporation, Government Agency, etc.)

(Your departmental administrator may be of assistance in identifying funding sources used)

Has external sponsor been informed of the invention?

yes

5. Internal Sponsor (Departmental Research Funds, etc.):

no

6. State first date of:

a. Conception 11/1/92

b. Sketch or drawing none

c. Written description 5/15/93

d. Completion of working model (or operational process) 10/1/94

7. State first date of:
- a. Disclosure to others (non-CMU employees)
_____ 4/4/95 _____
 - b. Printed publication _____ 7/8/95 _____
 - c. Oral disclosure (e.g. seminars, conferences, etc.) _____ 11/1/93 _____
 - d. Use for profit _____ none _____
 - e. Offer for sale _____ none _____

8. If your work has been disclosed to parties outside of Carnegie Mellon, please give a detailed explanation of the circumstances surrounding this external disclosure. Was it made under the provisions of a Funding Agreement or governed by a Confidentiality Agreement?

A pre-print of a research paper that will be published in the 7/8/95 Conference on Machine Learning has been available since 3/15/95 on the World-Wide Web. Many people have downloaded copies since then. There have been no confidentiality agreements made regarding the information concerning this work.

9. What makes this invention novel? How does the invention differ from present technology? What problems does it solve, or what advantages does it possess?

It is new in that some conventional Information Retrieval and Machine Learning techniques have been applied to this domain in ways not done previously. Present technology for filtering netnews tends to rely on the reader building a model of his or her interests by specifying keywords, whereas NewsWeeder tries to learn this information by watching what the user likes.

10. What are the present and future uses, applications and advantages of this invention:

The uses are for filtering information, documents, web pages, etc. without the user needing to specify keywords. However, it is more applicable to long-term interests in which many examples of documents the user likes are available, as opposed to a short-term information quest. NewsWeeder can also be used to categorize documents by learning from an already established set of such documents.

11. What are the disadvantages or limitations of this invention?

The research completed so far was the student's first foray into this field, and so the technology was created as more of a vehicle for research than a practical solution for more than a handful of researchers to actually use. Therefore it will need to be redeveloped completely as a technology to have any commercial value. Its main value at the moment is in demonstrating the skill of the inventor and as a general proof-of-concept that it at least can be done, if only for a few people. Some sections of the software code may have use in future projects, but since it is written entirely in the programming language PERL, it has been estimated to be about 100 times slower than if rewritten in an efficient languages like C, and so will probably not be a good choice.

12. Has the invention been tested experimentally? Are experimental data available?

While no experiments have been done on real users using the system, as part of the thesis research, Ken has taken data collected from the system's use and tried off-line experiments. The results from those experiments are available in the ML95 paper mentioned previously.

13. Give your opinion on the current stage of development of the invention as it relates to its current marketability

embryonic -> it will need a full team of programmers to recreate the software and Ken will be spending the second half of his thesis research efforts on developing new techniques to make the system more practical from the ground up
(needs substantial work to bring to market)

partially developed
(could be brought to market with nominal investment)

off-the-shelf
(could be brought to market with minimal investment)

14. Do you know of any other inventions which are related to this invention?

See the ML95 paper on the World-Wide Web for a list of related papers and work.
The paper can be retrieved at URL <http://anther.learning.cs.cmu.edu/ifhome/ml95.ps>

15. Has there been a prior art/patent search relating to this invention?
(please include copies of any resulting documentation)

No.

16. Has there been a literature search relating to this invention?
(please include copies of any resulting documentation)

Not a very complete one, although a bibliography will be prepared as a regular part of completion of Ken's thesis.

17. Has any intellectual property protection been applied to this invention? If so, please describe the specifics:

a. patent - No

b. copyright - Nothing official, but default copyright rules would apply to the software and research paper.

c. trademark - After seeing a strong positive reaction by his friends to the name Ken came up with for his project, "NewsWeeder", Ken has had the name trademarked as "intent-to-use" in case it becomes commercial someday.

18. Have you been personally involved in any prior patent processes for any other technology? If so, when, where, and for what type of invention?

No.

19. Has the invention been disclosed to industry representatives? Has any commercial interest been shown in it? Name companies and specific individuals and their titles.

After putting the ML95 research paper out on the Web, many people from academia and industry have expressed interest in knowing more about the project via email, asking questions, etc. One smaller company in particular, JLS, Inc., has also expressed an interest in helping Ken create his next research project, in exchange for licensing that new technology if it is commercially useful to them. They do not have any commercial interest in the technology created so far except maybe as a demo system for generating interest for the development of the next project. The contact person from JLS that has been spoken to is the president, Larry Sullivan. No help, agreements, or funding has been supplied by JLS at this time.

20. Give your opinion on potential marketability including possible commercial suggestions and the potential long-term commercial interest. Please specify any application areas and/or products you feel may embody aspects of your technology.

Some future incarnation of NewsWeeder could be very important to people needing information filtering services, if it can beat all the other major players trying to do the same thing in the long-term.

21. Please recommend three or four professionals from the CMU community who would be best suited to evaluate the technical and commercial merits of this technology:

Name: Tom Mitchell Dept.: SCS

Ext.: 8-2611

Name: Andy Witkin Dept.: SCS

Ext.: 8-6244

Name: Katia Sycara Dept.: RI

Ext.: 8-8825

Name: Dept.:

Ext.:

22. Do you know of other firms that might be particularly interested in the invention?

There is a general interest probably by any company that works with large amounts of text information, but most will want a more "fully-baked" type of technology before making any major commitments.

23. This invention has been witnessed and understood by me:

Ken Lang

print or type name signature date

print or type name signature date

Please feel free to attach additional material or data that would provide us with helpful information.

Please return to: Barb Mulholland, Technology Transfer Office, Warner Hall 405