

# COS 217 Final Exam

Princeton University, Spring 2005

May 13, 2005

This exam is **open book, open notes**. You have **3 hours**.

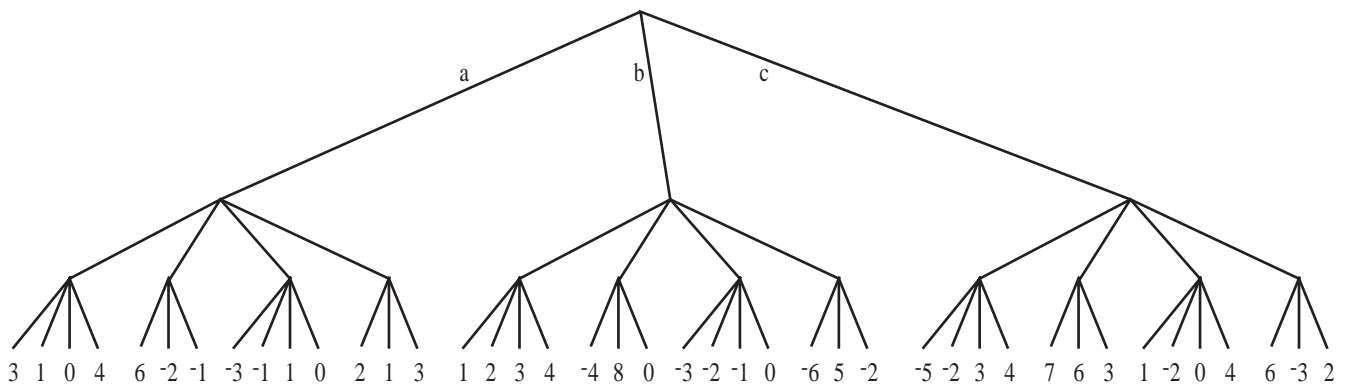
**Your name:**

**Write and sign the honor pledge:**

Below is a depth-3 game tree for a two-player game, where moves strictly alternate between the players. It's the MAX player's turn to move, and he wishes to maximize his score.

**Question 1 (2 points).** In a standard (non-alpha-beta) minimax search of depth 2, how many times will the heuristic function be evaluated?

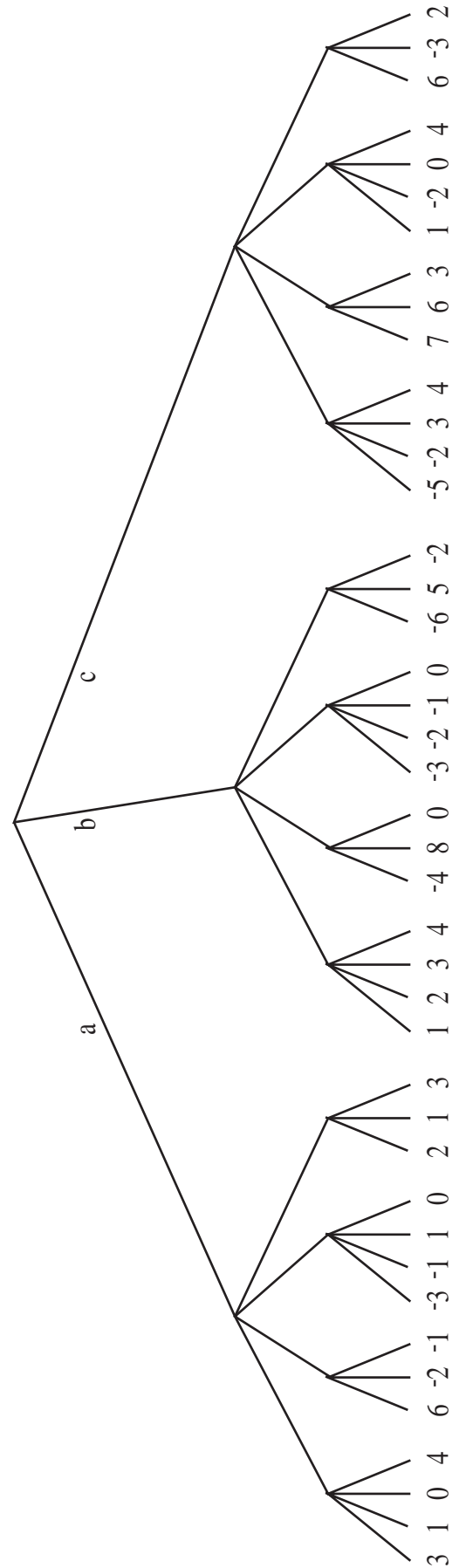
**Question 2 (3 points).** In a standard (non-alpha-beta) minimax search of depth 3, how many times will the heuristic function be evaluated? Which of the three moves will MAX choose? Show your work.



At right is a copy of the very same tree.

**Question 3 (11 points).** In an alpha-beta search of depth 3, how many times will the heuristic function be evaluated? Indicate with X's which subtrees will never be searched. Which of the three moves will MAX choose? Show your work.

**Question 4 (2 points).** Suppose player MAX is a Jedi knight, with the power to convince an opponent to make a particular choice, as illustrated in the original Star Wars movie (episode 4). What is MAX's best move?



```

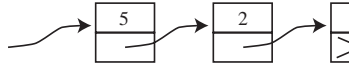
struct list {int val; struct list *next;};

int f (struct list *p, struct list *r) {
    struct list copy;
    if (p == NULL) {
        int sum = 0;
        for (; r != NULL; r = r->next)
            sum = 2*sum + r->val;
        return sum;
    }
    else {
        copy.val = p->val;
        copy.next = r;
        return f (p->next, &copy);
    }
}

```

```

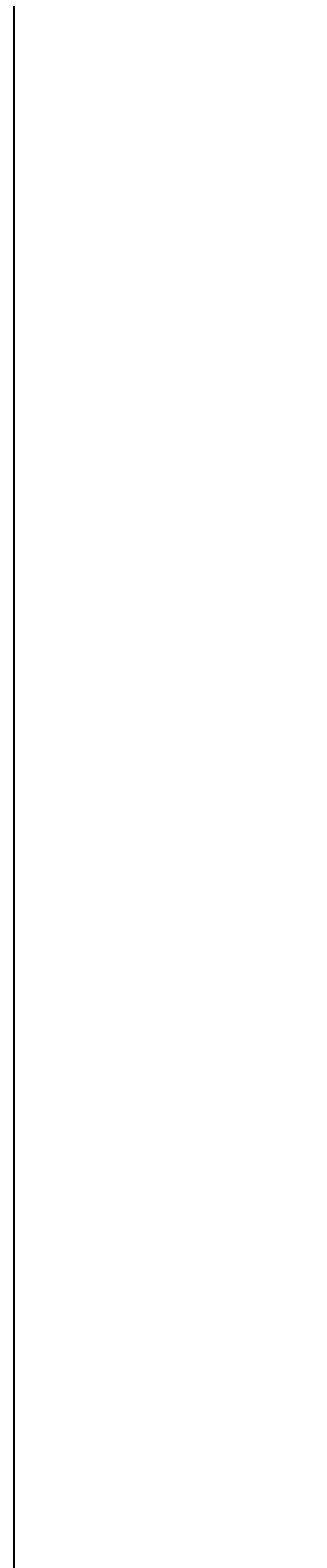
int main(int argc, char **argv) {
    struct list *mylist;

    mylist = 
    printf ("%d\n", f (mylist, NULL));
    return 0;
}

```

**Question 5 (15 points).** When the line `return sum;` is reached, draw a picture of the stack frames of `f` and `main`. Label the various parts. Your picture should indicate the *values of all the offsets* necessary for translating this program to assembly language, and the *contents of all the integer and pointer variables* in this execution of the program.

**Question 6 (10 points).** What number does this program print?



Bottom of Stack

**Question 7 (15 points).** Translate the function `f` to Intel IA-32 (x86, Pentium) Unix-style assembly language. Include comments to indicate the correspondence between assembly language instructions and C statements.

Here's the `symtable.h` you used in a programming assignment:

```
typedef struct SymTable_struct *SymTable_T;

SymTable_T SymTable_new(void);

void SymTable_free(SymTable_T oSymTable);

unsigned int SymTable_getLength(SymTable_T oSymTable);

int SymTable_put(SymTable_T oSymTable, const char *pcKey, const void *pvValue);

int SymTable_remove(SymTable_T oSymTable, const char *pcKey);

int SymTable_contains(SymTable_T oSymTable, const char *pcKey);

void *SymTable_get(SymTable_T oSymTable, const char *pcKey);

void SymTable_map(SymTable_T oSymTable,
    void (*pfApply)(const char *pcKey, void *pvValue, void *pvExtra),
    const void *pvExtra);
```

As you know, the keys of the symbol table can be arbitrary strings, not just “words”.

We wish to augment the interface to provide another operation, `SymTable_mapWords`, that's very much like `SymTable_map` except that it skips (doesn't call `pfApply` on) any symbol that doesn't start with an alphabetic character.

**Question 8 (20 points).** Add the declaration of `SymTable_mapWords` to the header file above, and show the implementation of such a function in `symtable.c` below. Don't bother with comments or assertions.

```
/* symtable.c */

#define BUCKETS 1017

struct binding {
    char *key;
    void *value;
    struct binding *next;
};

struct SymTable_struct {
    struct binding *bucket[BUCKETS];
};
```

```
/* symtable.c, continued ... */
static unsigned int hash(char *key) {
    int i;
    unsigned int h = 0U;
    for (i = 0; key[i] != '\0'; i++)
        h = h * 65599 + key[i];
    return h;
}

/* symtable.c, continued ... put your answer below */
```

In lecture I gave you this algorithm for computing  $p_{Spam}$ , the probability that this message is spam.

We say  $i \in \text{test}$  if feature  $i$  is in the test message.

$$p_0 = P(v_0) \cdot \prod_{i \in \text{test}} P(a_i|v_0) \cdot \prod_{i \notin \text{test}} 1 - P(a_i|v_0)$$

That is, to compute  $p_0$ , take the product of  $P(a_i|v_0)$  for just those features in the test message, and the product of  $(1 - P(a_i|v_0))$  for those features *not* in the test message.

$$p_1 = P(v_1) \cdot \prod_{i \in \text{test}} P(a_i|v_1) \cdot \prod_{i \notin \text{test}} 1 - P(a_i|v_1)$$

$$p_{Spam} = \frac{p_1}{p_0 + p_1} = \frac{1}{\frac{p_0}{p_1} + 1}$$

Now I propose a new method to compute  $p_0/p_1$ .

$$A = \frac{P(v_0)}{P(v_1)} \cdot \prod_i \frac{1 - P(a_i|v_0)}{1 - P(a_i|v_1)} \quad H_i = \frac{P(a_i|v_0)(1 - P(a_i|v_1))}{P(a_i|v_1)(1 - P(a_i|v_0))}$$

Notice that  $A$  and the  $H_i$  don't depend on what features are in the test message, so they can be computed in advance.

**Question 9 (5 points).** Prove that

$$\frac{p_0}{p_1} = A \cdot \prod_{i \in \text{test}} H_i$$

*Note: This answer may be tedious and time-consuming. Read and understand the question now, then come back and answer it later.*



The next page shows an e-mail I received last year, recognized by Spam Assassin as spam. In the middle you can see the extra mail headers that Spam Assassin inserted, such as **X-Spam-Report** and **X-Spam-Status**.

You can see that Spam Assassin works by assigning each feature a plus or minus “hit score,” and then it simply adds up the scores to see if they exceed a threshold. In this example,

**X-Spam-Status: Yes, hits=14.1 required=5.0**

the score added up to 14.1, exceeding the threshold of 5.0. Each user can set the threshold to his or her preference; after a few months at 5.0, I decided to lower my threshold to 3.0.

**Question 10 (7 points).** Spam Assassin has employees whose job is to determine appropriate values for the “hit scores”, such as a score of +2.8 for the feature SUBJ\_VIAGRA. What method should they use to set a “hit score” for each feature?

**Question 11 (2 points).** Suppose the hit scores are computed as you propose. I want SpamAssassin to reject a message exactly when its probability of being spam is 90%. What method should I use to set my threshold?

From EEKKAYBABJVXG@yahoo.com Wed Mar 24 07:51:34 2004  
Return-Path: <EEKKAYBABJVXG@yahoo.com>  
Received: from ribavirin.CS.Princeton.EDU (ribavirin.CS.Princeton.EDU [128.112.136.44])  
by memphis.CS.Princeton.EDU (8.12.11/8.12.11) with SMTP id i20CpUT7029354  
for <kappel@memphis.CS.Princeton.EDU>; Wed, 24 Mar 2004 07:51:34 -0500 (EST)  
Received: from bluebox.CS.Princeton.EDU ([128.112.136.38])  
by ribavirin.CS.Princeton.EDU (NAVGW 2.5.2.17) with SMTP id M2004032407513313096  
for <kappel@memphis.CS.Princeton.EDU>; Wed, 24 Mar 2004 07:51:33 -0500  
Received: from Princeton.EDU (postoffice02.Princeton.EDU [128.112.130.38])  
by bluebox.CS.Princeton.EDU (8.12.11/8.12.11) with ESMTP id i20CpX5Y015415  
for <kappel@cs.Princeton.EDU>; Wed, 24 Mar 2004 07:51:33 -0500 (EST)  
Received: from postoffice02.Princeton.EDU (localhost [127.0.0.1])  
by Princeton.EDU (8.12.9/8.12.9) with SMTP id i20CpP66001355;  
Wed, 24 Mar 2004 07:51:25 -0500 (EST)  
Received: from Princeton.EDU ([128.112.130.38])  
by postoffice02.Princeton.EDU (SAVSMTP 3.1.3.37) with SMTP id M2004032407512506588  
; Wed, 24 Mar 2004 07:51:25 -0500  
Received: from dt-vhf-eca9.adsl.wanadoo.nl (dt-vhf-eca9.adsl.wanadoo.nl [81.69.10.169])  
by Princeton.EDU (8.12.9/8.12.9) with SMTP id i20CpJRu001305;  
Wed, 24 Mar 2004 07:51:20 -0500 (EST)  
Received: from 61.225.131.24 by 81.69.10.169; Wed, 24 Mar 2004 14:46:23 +0200  
Message-ID: <QAWVWEJJNPGMZIMJASWBSWR@yahoo.com>  
X-Habeas-SWE-1: winter into spring  
X-Habeas-SWE-2: brightly anticipated  
X-Habeas-SWE-3: like Habeas SWE (tm)  
X-Habeas-SWE-4: Copyright 2002 Habeas (tm)  
X-Habeas-SWE-5: Sender Warranted Email (SWE) (tm). The sender of this  
X-Habeas-SWE-6: email in exchange for a license for this Habeas  
X-Habeas-SWE-7: warrant mark warrants that this is a Habeas Compliant  
X-Habeas-SWE-8: Message (HCM) and not spam. Please report use of this  
X-Habeas-SWE-9: mark in spam to <<http://www.habeas.com/report/>>.  
From: "Eduardo Cano" <EEKKAYBABJVXG@yahoo.com>  
Reply-To: "Eduardo Cano" <EEKKAYBABJVXG@yahoo.com>  
To: blake@Princeton.EDU  
Cc: hmorr@Princeton.EDU, murray@Princeton.EDU, awhite@Princeton.EDU,  
mflusser@Princeton.EDU, pholmes@Princeton.EDU, sadryer@Princeton.EDU,  
appel@Princeton.EDU, betty@Princeton.EDU, adm16@Princeton.EDU,  
cjohnson@Princeton.EDU, rwijetun@Princeton.EDU, jshiple@Princeton.EDU,  
mrnolta@Princeton.EDU  
Subject: get your viagra  
Date: Wed, 24 Mar 2004 17:44:23 +0500  
X-Mailer: Internet Mail Service (5.5.2650.21)  
MIME-Version: 1.0  
Content-Type: multipart/alternative;  
boundary="--86235228766990023568"  
X-Priority: 5  
X-MSMail-Priority: Low  
X-Spam-KB: <http://www.Princeton.EDU/spam>  
X-Spam-Flag: YES  
X-Spam-Checker-Version: SpamAssassin 2.63 (2004-01-11) on  
memphis.CS.Princeton.EDU  
X-Spam-Report:  
\* 2.8 SUBJ\_VIAGRA Subject includes "viagra"  
\* 0.1 MIME\_HTML\_ONLY BODY: Message only has text/html MIME parts  
\* 0.0 HTML\_MESSAGE BODY: HTML included in message  
\* 2.2 HTML\_IMAGE\_ONLY\_02 BODY: HTML: images with 0-200 bytes of words  
\* 0.7 MIME\_HTML\_NO\_CHARSET RAW: Message text in HTML without charset  
\* 0.5 FORGED\_YAHOO\_RCVD 'From' yahoo.com does not match 'Received' headers  
\* -8.0 HABEAS\_SWE Has Habeas warrant mark (<http://www.habeas.com/>)  
\* 1.1 MIME\_HTML\_ONLY\_MULTIMultipart message only has text/html MIME parts  
\* 0.1 MISSING\_OUTLOOK\_NAME Message looks like Outlook, but isn't  
\* 1.1 FORGED\_MUA\_IMS Forged mail pretending to be from IMS  
\* 1.2 MISSING\_MIMEOLE Message has X-MSMail-Priority, but no X-MimeOLE  
\* 4.3 FORGED\_IMS\_HTML IMS can't send HTML message only  
\* 2.0 LOCAL\_DRUGS\_MALEDYSFUNCTION LOCAL\_DRUGS\_MALEDYSFUNCTION  
\* 4.3 FORGED\_IMS\_TAGS IMS mailers can't send HTML in this format  
\* 1.7 HTML\_MIME\_NO\_HTML\_TAG HTML-only message, but there is no HTML tag  
X-Spam-Status: Yes, hits=14.1 required=5.0 tests=FORGED\_IMS\_HTML,  
FORGED\_IMS\_TAGS, FORGED\_MUA\_IMS, FORGED\_YAHOO\_RCVD, HABEAS\_SWE,  
HTML\_IMAGE\_ONLY\_02, HTML\_MESSAGE, HTML\_MIME\_NO\_HTML\_TAG,  
LOCAL\_DRUGS\_MALEDYSFUNCTION, MIME\_HTML\_NO\_CHARSET, MIME\_HTML\_ONLY,  
MIME\_HTML\_ONLY\_MULTIMISSING\_MIMEOLE, MISSING\_OUTLOOK\_NAME, SUBJ\_VIAGRA  
autolearn=no version=2.63  
X-Spam-Level: \*\*\*\*\*  
Status: RO  
X-Status:  
X-Keywords:  
X-UID: 388

---86235228766990023568  
Content-Type: text/html;  
Content-Transfer-Encoding: quoted-printable

<CENTER><A HREF=3D"http://www.milkywilky.com/vig/"><img src=3D"http://euro=  
ultra.com/vig/mailler11.gif"></A></CENTER><BR><BR><BR><BR><BR>  
<CENTER><A HREF=3D"http://www.milkywilky.com/vig/blurry/">Remove Me</A></C=  
ENTER>

---86235228766990023568---

**Question 12 (5 points).** Instead of using Spam Assassin, suppose you are using a Naive Bayesian spam filter that removes just those messages that it thinks are at least 90% likely to be spam, and this 90% threshold is *not adjustable*. Even so, as a user you would like to filter out more of the spam. What other parameters could you adjust in Naive Bayesian to get the same effect?

**Question 13 (3 points).** Read the headers labeled **X-Habeas-SWE**. Should we “brightly anticipate” this solution to the spam problem? Why or why not? Use evidence from this message to explain what’s going on.

**END OF EXAM. (13 questions, 100 total points.)**