

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

מבוא למדעי המחשב 67101

תרגיל 4 - מערכים ושימוש בקבצים

להגשה בתאריך 16/5/2018 בשעה 23:55

הקדמה

בתרגיל זה נתרגל שימוש בלולאות, רשימות רב מימדיות, מילונים, פרמטרים בשורת פקודה ועבודה עם קבצים. בנוסף, תתנסו **בעבודה בזוגות** (נהלי ההגשה מופיעים לאחר תיאור התרגיל).

מטרת התרגיל היא כתיבת תכנית אשר מוצאת מילים בתוך מטריצה של אותיות (תפזורת). התכנית מקבלת כפרמטרים בשורת הפקודה את שם קובץ המילים, המכיל רשימה של מילים תקינות אותן נרצה לחפש במטריצה, שם קובץ קלט המכיל מטריצה של אותיות (תפזורת, ראו תיאור בהמשך), שם קובץ פלט אליו נכתוב את המילים שנמצאו ומספר הפעמים שהופיעו, ופרמטר של כיוון אשר יגדיר לנו את אופן החיפוש במטריצה. את התכניות נפעיל ע"י קריאה לסקריפט משורת הפקודה.

דגשים לתרגיל:

- יש לכתוב כל חלק בקובץ נפרד (שמות הקבצים מופיעים בהמשך). את החלוקה הפנימית לפונקציות עליכם לבצע על פי שיקול דעתכם, לפי העקרונות שנלמדו בקורס. הדגש צריך להיות על קוד מודולרי (ללא כפל קוד), ברור וקריא. את החלוקה לפונקציות יש לנמק בקובץ ה-README. ניתן להוסיף קבצים נוספים אך יש לנמק את הבחירה בקובץ ה-README ולוודא שההרצה עובדת.
- כל הקבצים יוגשו בתוך קובץ ex5.zip. ניתן להניח שכל הקבצים המוגשים בקובץ זה יימצאו בתיקיה בעת הקריאה לסקריפט.
- ניקוד יינתן גם על יעילות הפתרונות. נסו לוודא שאינכם מבצעים בדיקות מיותרות.
- שימו לב שהרצה מוצלחת של התכנית לא אמורה לגרור הדפסה כלשהי.
- סגנון: הקפידו על תיעוד נאות ובחרו שמות משתנים משמעותיים. הקפידו להשתמש בקבועים (שמות משתנים באותיות גדולות) על פי הצורך.
- את התרגיל יש לבצע בזוגות. אין הכוונה לחלק את העבודה בין שני חברי הקבוצה, אלא לכתוב ביחד את התרגיל ולהתייעץ אחד עם השני לגבי מבנה התכנית ודרכי המימוש.

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

חלק א'

בחלק זה נכתוב את התכנית crossword.py.

תיאור הבעיה:

עבור מטריצה של אותיות, למשל המטריצה הבאה:

	0	1	2	3	4
0	a	p	p	l	e
1	a	g	o	d	o
2	n	n	e	r	t
3	g	a	t	a	c
4	m	i	c	s	r

נרצה לדעת אילו מילים מתוך רשימת המילים נמצאות במטריצה, וכמה פעמים הן הופיעו.

רשימת מילים לדוגמה:

dog, cat, ants, apple, cake, long, short, can, toe, poet, crop.

אנחנו מקבלים כקלט גם את כיוון החיפוש. נשים לב כי ישנם שמונה כיוונים לחיפוש במטריצה דו-מימדית:

חץ	סימון (אות)	כיוון
↑	u	למעלה (לאורך אותה עמודה, מהשורה האחרונה עד לשורה הראשונה)
↓	d	למטה (לאורך אותה עמודה, מהשורה הראשונה עד לשורה האחרונה)
→	r	ימינה (לאורך אותה שורה, מהעמודה הראשונה לעמודה האחרונה)
←	l	שמאלה (לאורך אותה שורה, מהעמודה האחרונה לעמודה הראשונה)
↗	w	אלכסון עולה ימינה
↖	x	אלכסון עולה שמאלה
↘	y	אלכסון יורד ימינה
↙	z	אלכסון יורד שמאלה

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

התכנית תבצע את החיפוש במטריצה על פי האות הניתנת כפרמטר בשורת הפקודה. התכנית יכולה לקבל גם מספר אותיות (למשל $w \times l$ זה קלט תקין כפרמטר) ללא חשיבות לסדר, ובמקרה כזה החיפוש יתבצע בכל הכיוונים שניתנו כקלט.

עבור המטריצה ורשימת המילים לעיל, נראה דוגמאות לפלט בהינתן כיוונים שונים:

עבור u:

toe,1

עבור d:

poet,1

עבור l:

cat,1

dog,1

עבור w:

cat,1

עבור wl או lw:

cat,2

dog,1

בנספח מצורפות דוגמאות מפורטות עבור כלל כיווני החיפוש.

את התכנית נפעיל משורת הפקודה בלינוקס על ידי השורה הבאה:

```
python3 crossword.py <word_file> <matrix_file> <output_file> <directions>
```

פרמטרים:

- קובץ המילים - יכיל רשימה של מילים. כל מילה נמצאת בשורה נפרדת. ראו קובץ מילים לדוגמה - word_list.txt.
- קובץ מטריצה (או קובץ תפזורת) - יכיל מטריצה של אותיות על פי הקידוד הבא: כל שורה במטריצה נמצאת בשורה נפרדת בקובץ, האותיות בשורה מופרדות ע"י פסיק. מצורף קובץ mat.txt עם קידוד של המטריצה לדוגמה.
- קובץ פלט - אם הקובץ לא קיים, נפתח קובץ חדש בשם זה בתיקה הנוכחית ואילו יוכנס פלט התכנית. אם קיים קובץ בשם זה בתיקה הנוכחית, נדרוס את התוכן שלו עם פלט התכנית.
- כיווני חיפוש - רצף של אותיות המייצג את כיווני החיפוש בהתאם לטבלה לעיל.

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

הנחות על הקלט:

- לא ניתן להניח שהקבצים (אם קיימים) נמצאים באותה תיקיה עם הסקריפט.
- לא ניתן להניח שהאותיות במטריצה או בקובץ המילים באותיות קטנות או גדולות. אם במטריצה נמצאת המילה cAT וברשימת המילים נמצאת המילה cat, התכנית כן אמורה לספור זאת כהופעה של המילה במטריצה.
- לא ניתן להניח שמילה תופיע רק פעם אחת באותו כיוון חיפוש, או בפרט פעם אחת באותה שורה/עמודה/אלכסון.
- לא ניתן להניח שהמטריצות ריבועיות.
- לא ניתן להניח שהמילים בקובץ המילים מסודרות בסדר אלפביתי.
- ניתן להניח שקבצי הקלט, אם קיימים, נמצאים בפורמט המתואר ובעלי הרשאות קריאה.
- ניתן להניח שכל מילה בקובץ המילים מופיעה רק פעם אחת.

פלט:

הפלט של החיפוש הוא רשימה של מילים מרשימת המילים (מסודרות בסדר אלפביתי) ומספר הפעמים שמילים אלו הופיעו במטריצת הקלט (המילה והמספר מופרדים בפסיק). כל צמד של מילה ומספר ההופעות שלה נמצא בשורה נפרדת. הפלט כולל רק מילים שמספר ההופעות שלהן במטריצת הקלט הוא לפחות 1. המילים יופיעו באותיות קטנות (ראו קובץ פלט לדוגמה - outfile_udlrwxyz.txt).

טיפול בקלט ופלט:

- במקרה שמספר הפרמטרים לא תקין (שונה מ-4) יש להדפיס הודעת שגיאה:
ERROR: invalid number of parameters. Please enter word_file matrix_file output_file directions.
- במקרה שקובץ המילים או קובץ המטריצה לא קיים, יש להדפיס הודעת שגיאה שמודיעה למשתמש איזה קובץ לא נמצא ולסיים את הריצה.
עבור הקריאה עם קובץ מילים בשם word_list.txt שאינו קיים:
ERROR: Word file word_list.txt does not exist.
- עבור הקריאה עם קובץ מטריצה בשם mat.txt שאינו קיים:
ERROR: Matrix file mat.txt does not exist.
- אם שניהם לא קיימים מספיק להדפיס את ההודעה עבור קובץ המילים בלבד.
- במקרה שקלט כיווני החיפוש לא מגדיר כיוון חיפוש תקין (על פי הטבלה), יש להדפיס הודעת שגיאה:
ERROR: invalid directions.
- אין משמעות להופעה של אותו כיוון מספר פעמים והתכנה אמורה להתעלם מכך (נחשב תקין).
- במקרה שלא נמצאה אף מילה, קובץ הפלט צריך להיות ריק.
- קבצים ריקים הם גם קלט תקין - אם המטריצה ריקה או שקובץ המילים ריק, קובץ הפלט יהיה ריק.

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

התכנית צריכה לכלול את החלקים הבאים:

1. בדיקת תקינות הקלט.
2. פתיחת קבצי הקלט וטעינת תוכנם למבני נתונים של פייתון.
3. הגדרת מבנה נתונים עבור הפלט (אנחנו ממליצים להשתמש במילון אבל ניתן להגדיר גם אובייקטים אחרים).
4. גוף התכנית: ביצוע החיפוש בכיוונים שהתקבלו כקלט, תוך עדכון מבנה הנתונים של הפלט:
 - a. עבור כל אחת מאותיות החיפוש שהתקבלו כקלט:
 - i. ניתן לחלץ מהמטריצה רשימה של מחרוזות אשר מרכיבות את תוכן המטריצה בכיוון הזה. למשל, רשימה של שורות המטריצה בכיוון שמאל לימין. לפי המטריצה בדוגמה נקבל שהאיבר הראשון ברשימה כזו הוא המחרוזת "apple".
 - ii. עבור כל אחד מהרצפים בכיוון החיפוש הנתון:
 1. עבור כל אחת מהמילים ברשימת המילים (האם באמת חייבים לבדוק את כל המילים?):
 - a. נחפש את המילה ונעדכן את מבנה הנתונים של הפלט.
5. כתיבת תוצאות החיפוש לקובץ הפלט לאחר מיון.

מומלץ להפריד חלק זה למספר פונקציות.

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

חלק ב'

בחלק זה נממש את התכנית crossword3d.py.

תכנית זו זהה לאופן פעולת crossword מבחינת המטרה, הקלט והפלט (כולל ההנחות), מלבד הנקודות הבאות:

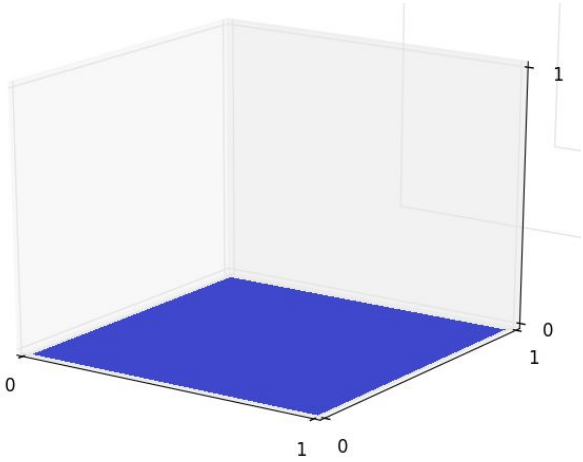
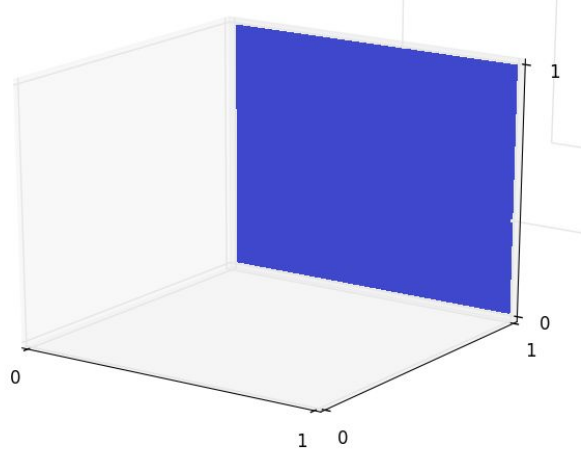
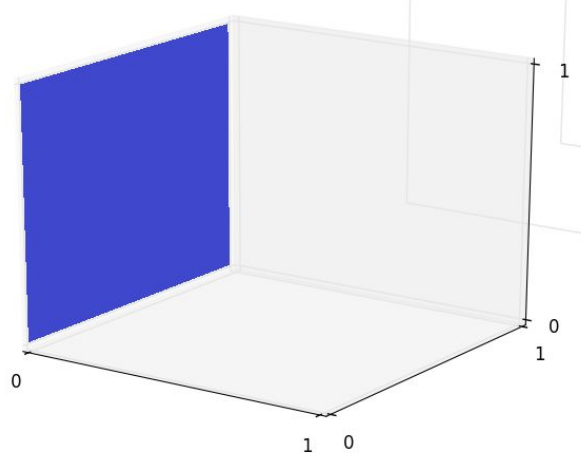
- קובץ התפזורת יכול כעת קידוד של מטריצה תלת מימדית. אפשר להסתכל על המטריצה התלת מימדית כמורכבת מרצף של מטריצות דו מימדיות (כמו שאפשר להסתכל על מטריצה דו מימדית כמורכבת מרצף של מטריצות חד מימדיות). כך למעשה הקובץ יכול מספר מטריצות דו מימדיות, לפי הפורמט שתואר בחלק א', כאשר בין כל שתי מטריצות דו מימדיות מופיעה שורה עם שלוש כוכביות (***) . בפרט, לאחר המטריצה הדו-מימדית האחרונה לא מופיעה שורת כוכביות. מטריצה תלת מימדית בעומק 1 תופיע ללא שורת כוכביות (זהה לקלט של מטריצה דו מימדית). ראו קובץ קלט לדוגמה - mat_3d.txt.
- כעת יש שלושה כיווני חיפוש המיוצגים ע"י אותיות אחרות - a, b, c (פירוט בטבלה בהמשך). בדומה לחלק א', צירוף של מספר אותיות הוא קלט תקין ומשמעותו שהפלט צריך לכלול את סכום המילים שנמצאו בכל כיווני חיפוש שניתנו.
- שימו לב כי בדיקת תקינות הקלט בחלק זה צריכה לוודא כי הכיוונים שהתקבלו הם חוקיים עבור כיווני החיפוש המוגדרים לחלק זה, ולא ביחס לכיוונים שהוגדרו בחלק א'. ניתן לחשוב על המטריצה התלת-מימדית כהרכבה של המטריצות הדו מימדיות, אחת על גבי השניה, כאשר הראשונה שמופיעה תהיה בעומק 0.

את התכנית נפעיל משורת הפקודה בלינוקס על ידי השורה הבאה:

```
python3 crossword3d.py word_file matrix_file output_file directions
```

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

בחלק זה קלט החיפוש כולל 3 כיוונים:

המחשה	סימון (אות)	כיוון
	a	<p>טבלאות עומק - חיפוש בכל הכיוונים שהוגדרו בחלק א' בכל טבלאות העומק (הטבלאות אשר מופרדות בכוכביות בקובץ התפזורת).</p> <p>בתמונה משמאל מסומנת טבלת העומק הראשונה. את החיפוש יש לבצע על כל טבלאות העומק.</p>
	b	<p>טבלאות אורך - חיפוש בכל הכיוונים שהוגדרו בחלק א' בכל טבלאות האורך.</p> <p>בתמונה משמאל מסומנת טבלת האורך הראשונה - מורכבת מהשורה הראשונה של כל אחת מטבלאות העומק. את החיפוש יש לבצע על כל טבלאות האורך. (ראו דוגמה לטבלאות אורך בקובץ length_matrix_2d.txt).</p>
	c	<p>טבלאות רוחב - חיפוש בכל הכיוונים שהוגדרו בחלק א' בכל טבלאות הרוחב.</p> <p>בתמונה משמאל מסומנת טבלת הרוחב הראשונה - מורכבת מהעמודה הראשונה של כל אחת מטבלאות העומק. את החיפוש יש לבצע על כל טבלאות האורך. (ראו דוגמה לטבלאות רוחב בקובץ width_matrix_2d.txt).</p>

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

נשים לב כי על מנת לחפש בכיוון מסוים, כל שעלינו לעשות זה לייצר את המטריצות הדו-מימדיות המתאימות ולהעביר אותן לחיפוש בכלל הכיוונים שהוגדרו בסעיף א'.

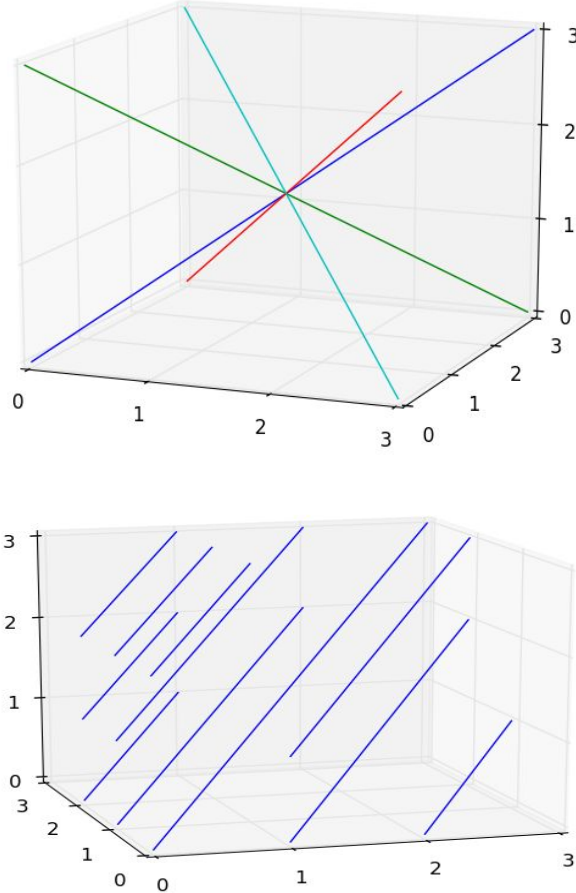
כלומר, התכנית צריכה לכלול את החלקים הבאים:

1. בדיקת תקינות הקלט.
2. פתיחת קבצי הקלט וטעינת תוכנם למבני נתונים של פייתון.
3. הגדרת מבנה נתונים עבור הפלט (אנחנו ממליצים להשתמש במילון אבל ניתן להגדיר גם אובייקטים אחרים).
4. גוף התכנית: ביצוע החיפוש בכיוונים שהתקבלו כקלט, תוך עדכון מבנה הנתונים של הפלט:
 - a. עבור כל אחת מאותיות החיפוש שהתקבלו כקלט:
 - i. נבנה רשימה של מטריצות דו מימדיות הרלוונטיות לכיוון זה.
 - ii. עבור כל מטריצה דו מימדית:
 1. עבור כל אחד מ-8 כיווני החיפוש שהוגדרו בסעיף א':
 - a. ניתן לחלץ מהמטריצה רשימה של מחרוזות אשר מרכיבות את תוכן המטריצה בכיוון הזה. למשל, רשימה של שורות המטריצה בכיוון שמאל לימין.
 - b. עבור כל אחד מהרצפים בכיוון החיפוש הנתון:
 - i. עבור כל אחת מהמילים במילון (האם באמת חייבים לבדוק את כל המילים?):
 1. נחפש את המילה ונעדכן את מבנה הנתונים של הפלט.
5. מומלץ להפריד חלק זה למספר פונקציות.
5. כתיבת תוצאות החיפוש לקובץ הפלט לאחר מיון.

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

סעיף בונוס

הוסיפו כיוון חיפוש נוסף אשר יסומן באות n:

המחשה	סימון (אות)	כיוון
	n	<p>אלכסונים - 8 כיווני החיפוש באלכסון בתוך המטריצה.</p> <p>בתמונה העליונה משמאל מופיעים ארבעת האלכסונים הראשיים של המטריצה. שימו לב שבכל ישר המסומן בתמונה ניתן לחפש בשני כיוונים, במעלה הישר ובמורד הישר.</p> <p>את החיפוש יש לבצע על כל האלכסונים: עבור כל כיוון חיפוש המוגדר ע"י אלכסון בתמונה העליונה, יש לחפש באלכסון היוצא בכיוון זה מכל משבצת במטריצה. עם זאת, נזכיר כי יש להימנע מבדיקות כפולות.</p> <p>בתמונה התחתונה משמאל מופיעים חלק מהאלכסונים במטריצה הנמצאים באותו כיוון חיפוש.</p>

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

הוראות הגשה

עליכם להגיש את הקובץ ex5.zip (בלבד) בקישור ההגשה של תרגיל 5 דרך אתר הקורס על ידי לחיצה על "Upload file". אנו ממליצים להתחיל לעבוד על התרגיל בשלב מוקדם שכן התרגיל ארוך. את התרגיל תבצעו בזוגות. הגשה בזוגות דורשת ראשית יצירת קבוצה. ברגע שאתם נכנסים למודל להגשת התרגיל תידרשו ליצור קבוצה לפני ההגשה. אחד מבני הזוג ייצור קבוצה ע"י הזנת שם בן הזוג השני (שימו לב להוסיף את השם בדיוק כפי שהוא מופיע במודל, כולל אותיות גדולות במקום הנכון). בן הזוג השני יראה שהוא בקבוצה עם הראשון ע"י שייכנס להגשה ויראה את שם בן הזוג. אנא בצעו את הוספת בן הזוג כבר בשלב בו התחלתם לעבוד (אין צורך לבצע הגשה בפועל כדי לרשום בן זוג). אנא וודאו כי אין בקבוצה יותר מ-2 תלמידים. על כל זוג יש להגיש קובץ אחד בלבד (בן הזוג השני יראה את הקובץ המוגש דרך מערכת המודל). שימו לב, בתרגיל זה, מלבד הקבצים הסטנדרטיים, עליכם להגיש גם קובץ נוסף בשם AUTHORS (ללא כל סיומת). קובץ זה יכיל שורה אחת ובו הלוגינים של שני הסטודנטים, מופרד ע"י פסיק. כך: minniemouse,mickeymouse

אנא וודאו כי הגשתם קובץ AUTHORS, אי הגשה שלו תגרור הורדה בציון.

ex5.zip צריך לכלול (לפחות) את הקבצים:

1. crossword.py
2. crossword3d.py
3. README - כולל את החלקים המפורטים בנהלי הקורס, הסברים על החלוקה הפנימית לפונקציות ולקבצים נוספים (אם יש). בתחילת הקובץ יופיעו הפרטים של שני המגישים מופרדים בפסיק.
4. AUTHORS

הנחיות כלליות בנוגע להגשה

- הנכם רשאים להגיש תרגילים דרך מערכת ההגשות באתר הקורס מספר רב של פעמים. ההגשה האחרונה בלבד היא זו שקובעת ושתיבדק.
- לאחר הגשת התרגיל, ניתן ומומלץ להוריד את התרגיל המוגש ולוודא כי הקבצים המוגשים הם אלו שהתכוונתם להגיש וכי הקוד עובד על פי ציפיותיכם.
- באחריותכם לוודא כי – PDF הבדיקות נראה כמו שצריך.
- קראו היטב את קובץ נהלי הקורס לגבי הנחיות נוספות להגשת התרגילים.
- שימו לב - יש להגיש את התרגילים בזמן!

בהצלחה!

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

נספח:

עבור הטבלה:

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

ורשימת המילים:

dog, cat, ants, apple, cake, long, short, can, toe, poet, crop.

הפלט עבור כלל כיווני החיפוש:

u:

toe,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

d:

poet,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

r:

apple,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

l:

cat,1

dog,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

w:

cat,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

בית הספר להנדסה ומדעי המחשב ע"ש רחל וסלים בנין

x:

can,1

crop,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

y:

ants,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r

z:

long,1

a	p	p	l	e
a	g	o	d	o
n	n	e	r	t
g	a	t	a	c
m	i	c	s	r