מטלת מנחה (ממ"ן) 14

הקורס: 20465 - מעבדה בתכנות מערכות

חומר הלימוד למטלה: פרויקט גמר

מספר השאלות: 1 | משקל המטלה: 31 נקודות

סמסטר: 2019א' | מועד אחרון להגשה: 10.3.2019

קיימות שתי חלופות להגשת מטלות:

- שליחת מטלות באמצעות מערכת המטלות המקוונת באתר הבית של הקורס
- שליחת מטלות באמצעות דואר אלקטרוני - באישור המנחה בלבד

הסבר מפורט ב"נוהל הגשת מטלות מנחה"

אחת המטרות העיקריות של הקורס "20465 - מעבדה בתכנות מערכות" היא לאפשר ללומדים בקורס להתנסות בכתיבת פרויקט תוכנה גדול, אשר יחקה את פעולתה של אחת מתוכניות המערכת השכיחות.

בפרויקט זה עליכם לכתוב תוכנת אסמבלר, עבור שפת אסמבלי שתוגדר בהמשך.
הפרויקט ייכתב בשפת C.

עליכם להגיש:

1. קבצי המקור של התוכנית שכתבתם (קבצים בעלי סיומת c.או h.).
2. קובץ הרצה (מקומפל ומקושר) עבור מערכת אובונטו.
3. קובץ makefile. יש להשתמש בקומפיילר gcc עם הדגלים: Wall –ansi –pedantic– ולנפות את ההודעות שמוציא הקומפיילר, כך שהתכנית תתקמפל ללא הערות או אזהרות.
4. דוגמאות הרצה (קלט ופלט):

א. קבצי קלט בשפת אסמבלי, וקבצי הפלט שנוצרו מהפעלת האסמבלר על קבצי קלט אלה. יש להדגים שימוש במגוון הפעולות וטיפוסי הנתונים של שפת האסמבלי.

ב. קבצי קלט בשפת אסמבלי המדגימים מגוון רחב של סוגי שגיאות אסמבלי (ולכן לא נוצרים קבצי פלט), ותדפיסי המסך המראים את הודעות השגיאה שמוציא האסמבלר.

בשל גודל הפרויקט, עליכם לחלק את התוכנית למספר קבצי מקור, לפי משימות. יש להקפיד שקוד התכנית יעמוד בקריטריונים של בהירות, קריאות וכתיבה נאה ומובנית.

נזכיר מספר היבטים חשובים:

1. הפשטה של מבני הנתונים: רצוי (במידת האפשר) להפריד בין הגישה למבני הנתונים לבין המימוש של מבני הנתונים. כך, למשל, בכתיבת שגרות לטיפול בטבלה, אין זה מעניינם של המשתמשים בשגרות אלה, האם הטבלה ממומשת באמצעות מערך או באמצעות רשימה מקושרת.

2. קריאות הקוד: יש להשתמש בשמות משמעותיים למשתנים ופונקציות. כמו כן, רצוי להגדיר קבועים רלוונטיים תוך שימוש בפקודת #define , ולהימנע משימוש ב-"מספרי קסם", שמשמעותם נהירה לכם בלבד. יש לערוך את הקוד באופן מסודר: הזחות עקביות, שורות ריקות להפרדה בין קטעי קוד, וכד'.

3. תיעוד: יש להכניס בקבצי המקור תיעוד תמציתי וברור, שיסביר את תפקידה של כל פונקציה ופונקציה. כמו כן יש להסביר את תפקידם של משתנים חשובים. כמו כן, יש להכניס הערות הסבר ברמת פירוט טובה בכל הקוד.

הערה: תוכנית "עובדת", דהיינו תוכנית שמבצעת את הדרוש ממנה, אינה ערובה לציון גבוה. כדי לקבל ציון גבוה על התכנית לעמוד בקריטריונים לעיל, אשר משקלם המשותף מגיע עד לכ-40% ממשקל הפרויקט.

מותר להשתמש בפרויקט בכל מגוון הספריות הסטנדרטיות של שפת C, אבל אין להשתמש בספריות חיצוניות אחרות.

מומלץ לעבוד בזוגות. אין לעבוד בצוותים גדולים יותר. **פרויקט שיוגש על ידי שלשה או יותר, לא ייבדק ולא יקבל ציון**. חובה שסטודנטים, הבוחרים להגיש יחד את הפרויקט, יהיו **שייכים לאותה קבוצת הנחיה**. הציון יהיה זהה לשני הסטודנטים.

מומלץ לקרוא את הגדרת הפרויקט פעם ראשונה ברצף, לקבלת תמונה כללית לגבי הנדרש, ורק לאחר מכן לקרוא בשנית, בצורה מעמיקה יותר.

רקע כללי ומטרת הפרויקט

כידוע, קיימות שפות תכנות רבות, ומספר גדול של תוכניות, הכתובות בשפות שונות, עשויות לרוץ באותו מחשב עצמו. כיצד "מכיר" המחשב כל כך הרבה שפות? התשובה פשוטה: המחשב מכיר למעשה שפה אחת בלבד: הוראות ונתונים הכתובים בקוד בינארי. קוד זה מאוחסן בגוש בזיכרון, ונראה כמו רצף של ספרות בינאריות. יחידת העיבוד המרכזית - היע"מ (CPU) - יודעת לפרק את הרצף הזה לקטעים קטנים בעלי משמעות: הוראות, מענים ונתונים. אופן הפירוק נקבע, באופן חד משמעי, על ידי המיקרו קוד של המעבד.

למעשה, זיכרון המחשב כולו הוא אוסף של סיביות, שנוהגים לראותן כמקובצות ליחידות בעלות אורך קבוע (בתים, מילים). לא ניתן להבחין, בעין שאינה מיומנת, בהבדל פיסי כלשהו בין אותו חלק בזיכרון שבו נמצאת תוכנית לבין שאר הזיכרון.

יחידת העיבוד המרכזית (היע"מ) יכולה לבצע מספר מסוים של הוראות פשוטות, ולשם כך היא משתמשת בזיכרון המחשב ובאוגרים (registers) הקיימים בתוך היע"מ. דוגמאות: העברת מספר מתא בזיכרון לאוגר ביע"מ או בחזרה, הוספת 1 למספר הנמצא באוגר, בדיקה האם מספר המאוחסן באוגר שווה לאפס. הוראות פשוטות אלה ושילובים שלהן הן המרכיבות את תוכנית המשתמש כפי שהיא נמצאת בזיכרון. כל תוכנית מקור (התוכנית כפי שנכתבה בידי המתכנת), תתורגם בסופו של דבר באמצעות תוכנה מיוחדת לצורה סופית זו.

קוד בשפת מכונה הוא רצף של ביטים המהווים קידוד של סדרת הוראות (תוכנית) שעל היע"מ לבצע. קוד מכונה אינו קריא למשתמש, ולכן לא נוח לקודד(או לקרוא) תכניות ישירות בשפת מכונה. שפת אסמבלי (assembly language) היא שפת תכנות מאפשרת לייצג את ההוראות של שפת המכונה בצורה סימבולית. כמובן שיש צורך לתרגם את הייצוג הסימבולי לקוד מכונה כדי שהתוכנית תוכל לרוץ במחשב. תרגום זה נעשה באמצעות כלי שנקרא אסמבלר (assembler).

כידוע, לכל שפת תכנות עילית יש מהדר (compiler) , או מפרש (interpreter), המתרגם תוכניות מקור לשפת מכונה. האסמבלר משמש בתפקיד דומה עבור שפת אסמבלי.

לכל מודל של יע"מ (אירגון של מחשב) יש שפת מכונה יעודית משלו, ובהתאם גם שפת אסמבלי יעודית משלו. לפיכך, גם האסמבלר (כלי התירגום) הוא יעודי ושונה לכל יע"מ.

תפקידו של האסמבלר הוא לייצר קוד מכונה גולמי עבור קובץ של תכנית הכתובה בשפת אסמבלי. זהו השלב הראשון במסלול אותו עוברת התכנית, עד לקבלת קוד המוכן לריצה על גבי חומרת המחשב. השלבים הבאים הם קישור (linkage) וטעינה (loading), אך בהם לא נעסוק בממ"ן זה.

המשימה בפרויקט זה היא לכתוב אסמבלר (כלומר תוכנית המתרגמת לשפת מכונה),עבור שפת אסמבלי שנגדיר כאן במיוחד לצורך הפרויקט.

לתשומת לב: בהסברים הכלליים על אופן עבודת תוכנת האסמבלר, תהיה מדי פעם התייחסות גם לעבודת שלבי הקישור והטעינה. התייחסויות אילו נועדו על מנת לאפשר לכם להבין את המשך תהליך העיבוד של הפלט של תוכנת האסמבלר. אין לטעות: עליכם לכתוב את תוכנית האסמבלר בלבד, **אין** צורך לכתוב גם את תוכניות הקישור והטעינה!!!

המחשב הדמיוני ושפת האסמבלי

נגדיר עתה את שפת האסמבלי ואת מודל המחשב הדמיוני, עבור פרויקט זה.

הערה: תאור מודל המחשב להלן הוא חלקי בלבד, ככל שנחוץ לביצוע המשימות בפרויקט.

"חומרה":

המחשב בפרויקט מורכב מיע"מ (יחידת עיבוד מרכזית), אוגרים וזיכרון RAM, כאשר חלק מהזיכרון משמש גם כמחסנית (stack). גודלה של מילת זיכרון במחשב הוא 12 סיביות.
האריתמטיקה נעשית בשיטת המשלים ל-2 (2's complement). מחשב זה עובד רק עם מספרים שלמים חיוביים ושליליים, אין תמיכה במספרים ממשיים.

אוגרים:

למעבד 8 אוגרים כלליים, בשמות r0, r1, r2, r3, r4, r5, r6, r7.
גודלו של כל אוגר הוא 12 סיביות. הסיבית הכי פחות משמעותית תצוין כסיבית מס' 0, והסיבית המשמעותית ביותר כמס' 11. שמות האוגרים נכתבים תמיד עם אות 'r' קטנה.

כמו כן יש במעבד אוגר בשם PSW (program status word), המכיל מספר דגלים המאפיינים את מצב הפעילות במעבד בכל רגע נתון. ראו בהמשך, בתאור הפקודות, לגבי השימוש בדגלים אלו.

גודל הזיכרון הוא 1024 תאים, בכתובות 0-1023 (בבסיס עשרוני), וכל תא הוא בגודל של 12 סיביות. לתא בזיכרון נקרא גם בשם "מילה". הסיביות בכל מילה ממוספרות בדומה לאוגר, כמפורט לעיל.

קידוד של תווים (characters) נעשה בקוד ascii.

מבנה הוראת מכונה:

כל הוראת מכונה מקודדת למספר מילות זיכרון, החל ממילה אחת ועד למקסימום שלוש מילים, הכל בהתאם לשיטות המיעון בהן נעשה שימוש (ראו בהמשך). בכל סוגי ההוראות, המבנה של המילה הראשונה זהה. מבנה המילה הראשונה בהוראה הוא כדלהלן:

11 9	8 5	4 2	1 0
מיעון אופרנד מקור	opcode	מיעון אופרנד יעד	A,R,E

סיביות 5-8 במילה הראשונה של ההוראה מהוות את קוד הפעולה (opcode). כל opcode מיוצג בשפת אסמבלי על ידי "שם פעולה". בשפה שלנו יש 16 קודי פעולה והם:

שם הפעולה	קוד הפעולה (בבסיס 10)
mov	0
cmp	1
add	2
sub	3
not	4
clr	5
lea	6
inc	7
dec	8

שם הפעולה	קוד הפעולה (בבסיס 10)
jmp	9
bne	10
red	11
prn	12
jsr	13
rts	14
stop	15

שמות הפעולות נכתבים תמיד באותיות קטנות. פרוט המשמעות של הפעולות יבוא בהמשך.

סיביות 0-1 (A,R,E)

סיביות אלה מציינות את סוג הקידוד, האם הוא מוחלט (Absolute) , חיצוני (External) או מצריך מיקום מחדש (Relocatable).

ערך של 00 משמעו שהקידוד הוא מוחלט.
ערך של 01 משמעו שהקידוד הוא חיצוני.
ערך של 10 משמעו שהקידוד מצריך מיקום מחדש.

סיביות אלה מתווספות רק לקידודים של הוראות (לא של נתונים), והן מתווספות גם לכל המילים הנוספות שיש לקידודים אלה.

ראו הסבר נוסף על סיביות אלה בהמשך.

סיביות 2-4 מקודדות את מספרה של שיטת המיעון של אופרנד היעד (destination operand).

סיביות 9-11 מקודדות את מספרה של שיטת המיעון של אופרנד המקור (source operand) .

<u>שיטות מיעון:</u>

בשפה שלנו קיימות שלוש שיטות מיעון, שמסומנות במספרים 1, 3, 5.

השימוש בשיטות מיעון מצריך קידוד של מילות-מידע נוספות בקוד המכונה של ההוראה.
אם בפקודה יש אופרנד אחד, תהיה מילת מידע אחת נוספת. אם בפקודה יש שני אופרנדים, ייתכנו שתי מילות-מידע נוספות, או מילת-מידע אחת המשותפת לשני האופרנדים, תלוי בשיטות המיעון בהן נעשה שימוש (ראו מפרט בהמשך). כאשר בקידוד הפקודה יש שתי מילות-מידע נוספות, אזי מילת-המידע הראשונה מתייחסת לאופרנד המקור, והשניה מתייחסת לאופרנד היעד.

בכל מילת-מידע נוספת, סיביות 0-1 הן השדה A,R,E.

להלן תיאור של שיטות המיעון:

קוד	שיטת מיעון	תוכן המילים הנוספות	אופן הכתיבה	דוגמא
1	מיעון מידי	מילת-מידע נוספת של ההוראה מכילה את האופרנד עצמו, שהוא מספר המיוצג ב- 10 סיביות, אליהן מתווספות זוג סיביות של השדה A,R,E.	האופרנד הוא מספר שלם בבסיס עשרוני	mov -103,@r2 בדוגמה זו האופרנד הראשון של הפקודה נתון בשיטת מיעון מיידי. ההוראה כותבת את הערך 103- אל אוגר r2
3	מיעון ישיר	מילת-מידע נוספת של ההוראה מכילה כתובת של מילה בזיכרון. מילה זו בזיכרון היא האופרנד. המען מיוצג ב- 10 סיביות אליהן מתווספות זוג סיביות של השדה A,R,E.	האופרנד הוא <u>תווית</u> שכבר הוצהרה או שתוצהר בהמשך הקובץ. ההצהרה נעשית על ידי כתיבת תווית בתחילת הנחיית 'data.'או 'string.', או בתחילת הוראה של התוכנית, או באמצעות אופרנד של הנחיית 'extern.'	נתונה למשל ההגדרה: x: .data 23 אפשר לכתוב הוראה: dec x בדוגמה זו, ההוראה מקטינה ב-1 את תוכן המילה שבכתובת x בזיכרון (ה"משתנה" x).
5	מיעון אוגר ישיר	האופרנד הוא אוגר. אם האוגר משמש כאופרנד יעד, מילת-מידע נוספת של הפקודה תקודד בחמש הסיביות 6-2 את מספרו של האוגר. ואילו אם האוגר משמש כאופרנד מקור, מספר האוגר יקודד בחמש הסיביות 11-7 של מילת-המידע. אם בפקודה יש שני אופרנדים, ושניהם אוגרים, הם יחלקו מילת-מידע אחת משותפת, כאשר הסיביות 6-2 הן עבור אוגר היעד, והסיביות 11-7 עבור אוגר המקור. בכל מילה נוספת מתווספות זוג סיביות של השדה A,R,E. סיביות שאינן בשימוש יכילו 0.	האופרנד הוא שם של אוגר. שם האוגר ייכתב כאשר לפניו יופיע התו @	mov @r1,@r2 בדוגמה זו, ההוראה מעתיקה את תוכן אוגר r1 אל אוגר r2. בדוגמה זו שני האופרנדים הם אוגרים, אשר יקודדו למילת-מידע נוספת אחת משותפת.

<u>הערה</u>: מותר להתייחס לתווית עוד לפני שמצהירים עליה, בתנאי שהיא אכן מוצהרת במקום כלשהו בקובץ.

אפיון הוראות המכונה:
הוראות המכונה מתחלקות לשלוש קבוצות, לפי מספר האופרנדים הדרוש להן.

קבוצת ההוראות הראשונה:
הוראות הדורשות שני אופרנדים. ההוראות השייכות לקבוצה זו הן:
mov, cmp, add, sub, lea

הפעולה	הסבר הפעולה	דוגמא	הסבר הדוגמא
mov	מבצעת העתקה של האופרנד הראשון, אופרנד המקור (source) אל האופרנד השני, אופרנד היעד (destination) (בהתאם לשיטת המיעון).	mov A,@r1	העתק את תוכן המשתנה A (המילה שבכתובת A בזכרון) אל אוגר r1.
cmp	מבצעת "השוואה" בין שני האופרנדים שלה. אופן ההשוואה: תוכן אופרנד היעד (השני) מופחת מתוכן אופרנד המקור (הראשון), ללא שמירת תוצאת החיסור. פעולת החיסור מעדכנת דגל בשם Z ("דגל האפס") באוגר הסטטוס(PSW).	cmp A, @r1	אם תוכן המשתנה A זהה לתוכנו של אוגר r1 אזי דגל האפס, Z, באוגר הסטטוס (PSW) יודלק, אחרת הדגל יאופס.
add	אופרנד היעד (השני) מקבל את תוצאת החיבור של אופרנד המקור (הראשון) והיעד (השני).	add A,@r0	אוגר r0 מקבל את תוצאת החיבור של תוכן המשתנה A ותוכנו הנוכחי של אוגר r0.
sub	אופרנד היעד (השני) מקבל את תוצאת החיסור של אופרנד המקור (הראשון) מאופרנד היעד (השני)	sub 3,@r1	אוגר r1 מקבל את תוצאת החיסור של הערך 3 מתוכנו הנוכחי של האוגר r1.
lea	lea הוא קיצור (ראשי תיבות) של: load effective address. פעולה זו מציבה את המען בזיכרון המיוצג על ידי התווית שבאופרנד הראשון (המקור), אל אופרנד היעד (האופרנד השני).	lea HELLO, @r1	המען שמייצגת התווית HELLO מוצב לאוגר r1.

קבוצת ההוראות השניה:
אלו הוראות הדורשות אופרנד אחד בלבד. אופן הקידוד של האופרנד הוא כמו של אופרנד היעד בפקודה עם שני אופרנדים. במקרה זה, השדה של אופרנד המקור (סיביות 4-2) במילה הראשונה בקידוד ההוראה הינו חסר משמעות, ולפיכך יכיל 0. על קבוצה זו נמנות ההוראות הבאות:
not, clr, inc, dec, jmp, bne, red, prn, jsr

פקודה	הסבר פעולה	דוגמא	הסבר דוגמא
not	היפוך ערכי הסיביות באופרנד (כל סיבית שערכה 0 תהפוך ל-1 ולהיפך: 1 ל-0).	not @r2	r2 ← not r2
clr	איפוס תוכן האופרנד.	clr @r2	r2 ← 0
inc	הגדלת תוכן האופרנד באחד.	inc @r2	r2 ← r2 + 1
dec	הקטנת תוכן האופרנד באחד.	dec C	C ← C – 1

פקודה	הסבר פעולה	דוגמא	הסבר דוגמא
jmp	קפיצה (הסתעפות) בלתי מותנית אל ההוראה שנמצאת במען המיוצג על ידי האופרנד. כלומר, בעת ביצוע ההוראה, מצביע התוכנית (PC) יקבל את ערך אופרנד היעד.	jmp LINE	PC ← LINE מצביע התכנית מקבל את המען המיוצג על ידי התווית LINE, ולפיכך הפקודה הבאה שתתבצע תהיה במען זה.
bne	bne הוא קיצור (ראשי תיבות) של branch if not equal (to zero). זוהי הוראת הסתעפות מותנית. מצביע התוכנית (PC) יקבל את ערך אופרנד היעד אם ערכו של הדגל Z באוגר הסטטוס (PSW) הינו 0. כזכור, ערך הדגל Z נקבע בהוראת cmp.	bne LINE	אם ערך הדגל Z באוגר הסטטוס (PSW) הינו 0: PC ← LINE
red	קריאה של תו מהקלט הסטנדרטי (stdin) אל האופרנד.	red @r1	קוד ה-ascii של התו הנקרא מהקלט הסטנדרטי ייכנס לאוגרr1.
prn	הדפסת התו הנמצא באופרנד, אל הפלט הסטנדרטי (stdout).	prn @r1	התו אשר קוד ה-ascii שלו נמצא באוגר r1 יודפס לפלט הסטנדרטי.
jsr	קריאה לשגרה (סברוטינה), מצביע התוכנית (PC) הנוכחי נדחף לתוך המחסנית שבזכרון המחשב, והאופרנד מוכנס ל-PC.	jsr FUNC	push(PC) PC ← FUNC

קבוצת ההוראות השלישית:
אלו הוראות ללא אופרנדים. קידוד ההוראה מורכב ממילה אחת בלבד. במקרה זה, השדות של אופרנד המקור ואופרנד היעד אינם רלוונטים (כי אין אופרנדים), ויכילו 0.

ההוראות השייכות לקבוצה זו הן : rts, stop.

פקודה	הסבר פעולה	דוגמא	הסבר דוגמא
rts	חזרה משיגרה. הערך בראש המחסנית של זמן-ריצה מוצא מן המחסנית ומוכנס אל מצביע התוכנית (PC).	rts	PC ← pop()
stop	עצירת ריצת התוכנית.	stop	התכנית עוצרת

מספר נקודות נוספות לגבי תיאור שפת האסמבלי:

שפת האסמבלי מורכבת ממשפטים (statements) כאשר התו המפריד בין משפט למשפט הינו תו '\n' (שורה חדשה). כלומר, כאשר מסתכלים על קובץ המקור, רואים אותו כמורכב משורות של משפטים, כאשר כל משפט מופיע בשורה נפרדת.

אורכה של שורה בקובץ המקור הוא 80 תווים לכל היותר (לא כולל התו \n).

ישנם ארבעה סוגי משפטים (שורות) בשפת האסמבלי, והם:

סוג המשפט	הסבר כללי
משפט ריק	זוהי שורה המכילה אך ורק תווים לבנים (whitespace), כלומר מכילה רק תווים '\t' ו-' ' (סימני tab ורווח). ייתכן ובשורה אין אף תו (למעט התו \n).
משפט הערה	זוהי שורה בה התו הראשון הינו ';' (נקודה פסיק). על האסמבלר להתעלם לחלוטין משורה זו.
משפט הנחיה	זהו משפט המנחה את האסמבלר מה עליו לעשות כשהוא פועל על תכנית המקור. יש מספר סוגים של משפטי הנחיה. משפט הנחיה עשוי לגרום להקצאה ואתחול משתנים של התכנית, אך הוא אינו מייצר קוד המיועד לביצוע בעת ריצת התכנית.
משפט הוראה	זהו משפט המייצר קוד לביצוע בעת ריצת התכנית. המשפט מכיל שם של פעולה שעל ה-CPU לבצע, ותיאור האופרנדים המשתתפים בפעולה.

כעת נפרט לגבי סוגי המשפטים השונים.

משפט הנחיה:

משפט הנחיה הוא בעל המבנה הבא:

בתחילתו יכולה להופיע תווית (label) (התווית חייבת להיות בתחביר חוקי. התחביר של תווית חוקית יתואר בהמשך). התווית היא אופציונלית.

לאחר מכן מופיע התו '.' (נקודה), ובצמוד לנקודה (ללא רווח) שם ההנחיה, באותיות קטנות (lower case) בלבד. לאחר שם ההנחיה יופיעו פרמטרים (מספר הפרמטרים נקבע בהתאם לסוג ההנחיה).

יש לשים לב: למילות הקוד הנוצרות ממשפט הנחיה לא מצורפות זוג סיביות A,R,E והקידוד ממלא את כל 12 הסיביות של המילה.

יש ארבעה סוגים של משפטי הנחיה , והם:

1. '.data'

הפרמטר(ים) של data.הם רשימת מספרים שלמים חוקיים (אחד או יותר) המופרדים על ידי התו ',' (פסיק). למשל:

.data 7, −57, +17, 9

יש לשים לב שהפסיקים אינם חייבים להיות צמודים למספרים. בין מספר לפסיק ובין פסיק למספר יכולים להופיע רווחים וטאבים בכל כמות (או בכלל לא), אולם הפסיק חייב להופיע בין המספרים. כמו כן, אסור שיופיע יותר מפסיק אחד בין שני מספרים, ולא פסיק אחרי המספר האחרון או לפני המספר הראשון.

משפט ההנחיה: '.data' מנחה את האסמבלר להקצות מקום בתמונת הנתונים (data image), אשר בו יאוחסנו הערכים המתאימים, ולקדם את מונה הנתונים, בהתאם למספר הערכים ברשימה. אם להוראת data. יש גם תווית, אזי תווית זו מקבלת את ערך מונה הנתונים (לפני הקידום), ומוכנסת אל טבלת הסמלים. דבר זה מאפשר להתייחס אל מקום מסוים בתמונת הנתונים דרך שם התווית (למעשה, זוהי הגדרת שם של משתנה).

כלומר אם נכתוב:

```
XYZ: .data 7, –57, +17, 9
```

אזי יוקצו בתמונת הנתונים ארבע מילים רצופות שיכילו את המספרים שמופיעים בהנחיה. התווית XYZ מזוהה עם כתובת המילה הראשונה.

אם נכתוב בתכנית הוראה לביצוע:

```
mov  XYZ, @r1
```

אזי בזמן ריצת התכנית יוכנס לאוגר r1 הערך 7.

ואילו ההוראה:

```
lea XYZ,@r1
```

תכניס לאוגר r1 את ערך התווית XYZ (כלומר הכתובת בזיכרון בה מאוחסן הערך 7).

2. '.string'

להנחיה '.string' פרמטר אחד שהוא מחרוזת חוקית. תווי המחרוזת מקודדים לפי ערכי ה-ascii המתאימים ומוכנסים אל תמונת הנתונים, לפי סדרם, כל תו במילה נפרדת. בסוף המחרוזת יוכנס הערך אפס, לסמן סיום מחרוזת. מונה הנתונים של האסמבלר יוגדל בהתאם לאורך המחרוזת + 1. אם בשורת ההנחיה מגדרת גם תווית, אזי ערכה יהיה מען המקום בזיכרון, שבו מאוחסן התו הראשון במחרוזת, באופן דומה כפי שנעשה עבור '.data'.

לדוגמא, משפט ההנחיה:

```
STR:  .string "abcdef"
```

מקצה בתמונת הנתונים "מערך תווים" באורך של 7 מקומות החל מהמען המזוהה עם התווית ABC, ומאתחל "מערך" זה לערכי ה-ascii של התווים: a, b, c, d, e, f בהתאמה, ולאחריהם הערך 0 לסימון סוף מחרוזת. התווית STR מזוהה עם כתובת התחלת המחרוזת.

3. '.entry'

להנחיה '.entry' פרמטר אחד והוא שם של תווית המוגדרת בקובץ המקור הנוכחי (כלומר מקבלת את ערכה בקובץ זה). מטרת entry. היא להצהיר על התווית הזו באופן שיאפשר לקטעי אסמבלי הנמצאים בקבצי מקור אחרים להשתמש בה.

לדוגמא, השורות:

```
        .entry   HELLO
HELLO:  add      1, @r1
        ........
```

מודיעות שאפשר להתייחס מקובץ אחר לתווית HELLO המוגדרת בקובץ הנוכחי.

לתשומת לב: תווית המוגדרת בתחילת שורת entry. הינה חסרת משמעות והאסמבלר מתעלם מתווית זו (אפשר שהאסמבלר יוציא הודעת אזהרה).

4. '.extern'

להנחיה '.extern' פרמטר אחד והוא שם של תווית. מטרת ההוראה היא להצהיר כי התווית מוגדרת בקובץ אחר וכי קטע האסמבלי, בקובץ זה, עושה בו שימוש. זוהי הנחיה תואמת להנחיית entry המופיעה בקובץ בו מוגדרת התווית. בזמן הקישור (link) תתבצע ההתאמה, בין ערך התווית, כפי שנקבע בקוד המכונה של הקובץ שהגדיר את התווית, לבין קוד ההוראות המשתמשות בו בקבצים אחרים.

לדוגמא, משפט ההנחיה '.extern' התואם למשפט ההנחיה '.entry' בדוגמא הקודמת יהיה:

```
.extern    HELLO
```

לתשומת לב: תווית המוגדרת בתחילת שורת extern. הינה חסרת משמעות והאסמבלר מתעלם מתווית זו (אפשר שהאסמבלר יוציא הודעת אזהרה).

משפט הוראה:

משפט הוראה מורכב מ:
1. תווית אופציונלית.
2. שם פעולה.
3. 0, 1 או 2 אופרנדים בהתאם לסוג הפעולה.

שם הפעולה נכתב תמיד באותיות קטנות (lower case), והפעולה היא אחת מבין 16 הפעולות שהוזכרו לעיל.

לאחר שם הפעולה יכולים להופיע אופרנדים (אחד או שניים).

כאשר יש שני אופרנדים, האופרנדים מופרדים זה מזה בתו ',' (פסיק). כמו בהנחה data., **לא חייבת להיות שום הצמדה של האופרנדים לפסיק או לשם הפעולה באופן כלשהו. כל כמות רווחים או tabs בין האופרנדים לפסיק ובין שם הפעולה לאופרנד הראשון היא חוקית.**

להוראה בעלת שני אופרנדים המבנה הבא:

```
תווית-אופציונלית:    שם-הפעולה    אופרנד-יעד, אופרנד-מקור
```

לדוגמא:

```
HELLO:    add    @r7, B
```

לפקודה בעלת אופרנד אחד המבנה הבא:

```
תווית-אופציונלית:    שם-הפעולה    אופרנד
```

לדוגמא:

```
HELLO:    bne    XYZ
```

להוראה ללא אופרנדים המבנה הבא:

```
תווית-אופציונלית:    שם-הפעולה
```

לדוגמא:

```
END:    stop
```

אם מופיעה תווית בשורת ההוראה, אזי היא תוכנס אל טבלת הסמלים. ערך התווית יהיה מען המילה הראשונה של ההוראה בתוך תמונת הקוד שבונה האסמבלר.

תווית:

תווית חוקית מתחילה באות אלפבית (גדולה או קטנה) ולאחריה סדרה כלשהי של אותיות אלפבית (גדולות או קטנות) וספרות. האורך המקסימלי של תווית הוא 31 תווים. הגדרת התווית מסתיימת על ידי התו ':' (נקודתיים). תו זה אינו מהווה חלק מהתווית. זהו רק סימן המציין את סופה. הגדרת התווית חייבת להתחיל בעמודה הראשונה בשורה. אסור שאותה תווית תוגדר יותר מפעם אחת (כמובן בשורות שונות). אותיות קטנות וגדולות נחשבות שונות זו מזו.

לדוגמה, התוויות המוגדרות להלן הן תוויות חוקיות.

hEllo:
x:
He78902:

לתשומת לב: מילים שמורות של שפת האסמבלי (כלומר שם של פעולה או הנחיה, או שם של אוגר) אינן יכולות לשמש כשם של תווית.

התווית מקבלת את ערכה בהתאם להקשר בו היא מופיעה. תווית בהנחיות data., string., תקבל את ערך מונה הנתונים (data counter) הנוכחי, בעוד שתווית המוגדרת בשורת הוראה תקבל את ערך מונה ההוראות (instruction counter) הנוכחי.

מספר:

מספר חוקי מתחיל בסימן אופציונלי '-' או '+' ולאחריו סדרה כלשהי של ספרות בבסיס עשר. הערך של המספר הוא הערך המיוצג על ידי מחרוזת הספרות והסימן. כך למשל, 76, 5-, 123+ הינם מספרים חוקיים. אין תמיכה בשלמים בייצוג אחר מאשר עשרוני, ואין תמיכה במספרים ממשיים.

מחרוזת:

מחרוזת חוקית היא סדרת תווי ascii נראים, המוקפים במרכאות כפולות (המרכאות אינן נחשבות כחלק מהמחרוזת). דוגמא למחרוזת חוקית: "hello world".

איפיון השדה A,R,E בקוד המכונה

בכל מילה בקוד המכונה של הוראה (לא של נתונים), האסמבלר מכניס מידע עבור תהליך הקישור והטעינה. זהו השדה A,R,E (שתי הסיביות הימניות 1-0). המידע ישמש לתיקונים בקוד בכל פעם שייטען לזיכרון לצורך הרצה. האסמבלר בונה מלכתחילה קוד שמיועד לטעינה החל מכתובת 100. התיקונים יאפשרו לטעון את הקוד בכל פעם למקום אחר, בלי צורך לחזור על תהליך האסמבלי.

שתי הסיביות בשדה A,R,E יכילו את אחד הערכים הבינאריים: 00, 10, או 01. המשמעות של כל ערך מפורטת להלן.

האות 'A' (קיצור של absolute) באה לציין שתוכן המילה אינו תלוי במקום בזיכרון בו ייטען בפועל קוד המכונה של התכנית בעת ביצועה (למשל מילה המכילה אופרנד מיידי).
במקרה זה שתי הסיביות הימניות יכילו את הערך 00.

האות 'R' (קיצור של relocatable) באה לציין שתוכן המילה תלוי במקום בזיכרון בו ייטען בפועל קוד המכונה של התכנית בעת ביצועה (למשל מילה המכילה כתובת של תווית המוגדרת בקובץ המקור). במקרה זה שתי הסיביות הימניות יכילו את הערך 10.

האות 'E' (קיצור של external) באה לציין שתוכן המילה תלוי בערכו של סמל חיצוני (external) (למשל מילה המכילה כתובת של תווית חיצונית, כלומר תווית שאינה מוגדרת בקובץ המקור).
במקרה זה שתי הסיביות הימניות יכילו את הערך 01.

אסמבלר עם שני מעברים

כאשר מקבל האסמבלר קוד לתרגום, עליו לבצע שתי משימות עיקריות לצורך הכנת הקוד הבינארי: הראשונה היא לזהות ולתרגם את שמות הפעולות, והשנייה היא לקבוע מענים לכל הסמלים המופיעים בתוכנית.

לדוגמא: האסמבלר קורא את קטע הקוד הבא:

```
MAIN:     mov    @r3 ,LENGTH
LOOP:     jmp    L1
          prn    -5
          bne    LOOP
          sub    @r1, @r4
          bne    END
L1:       inc    K
          bne    LOOP
END:      stop
STR:      .string "abcdef"
LENGTH:   .data  6,-9,15
K:        .data  22
```

עליו להחליף את שמות הפעולות mov, jmp, prn, sub, inc, bne, stop בקוד הבינארי השקול להם במודל המחשב שהגדרנו.

כמו כן, על האסמבלר להחליף את הסמלים K,STR, LENGTH, L1, MAIN, LOOP, END במענים של המקומות בזיכרון שם נמצאים כל נתון או הוראה בהתאמה.

נניח שקטע הקוד לעיל (הוראות ונתונים) ייטען בזיכרון החל ממען 100 (בבסיס 10) . במקרה זה נקבל את ה"תרגום" הבא:

Decimal Address	Source Code	Explanation	Binary Machine Code
0100 0101 0102	MAIN: mov @r3 ,LENGTH	first word of instruction source register 3 address of label LENGTH	**101000001100** **000110000000** **000111110110**
0103 0104	LOOP: jmp L1	 address of label L1	**000100101100** **000111000110**
0105 0106	prn -5	 immediate value -5	**000110000100** **111111101100**
0107 0108	bne LOOP	 address of label LOOP	**000101001100** **000110011110**
0109 0110	sub @r1, @r4	 registers r1 and r4	**101001110100** **000010010000**
0111 0112	bne END	 address of label END	**000101001100** **000111010110**
0113 0114	L1: inc K	 address of label K	**000011101100** **001000000010**
0115 0116	bne LOOP	 address of label LOOP	**000101001100** **000110011110**
0117	END: stop		**000111100000**

Decimal Address	Source Code	Explanation	Binary Machine Code
0118 0119 0120 0121 0122 0123 0124	STR: .string "abcdef"	ascii code 'a' ascii code 'b' ascii code 'c' ascii code 'd' ascii code 'e' ascii code 'f' ascii code '\0' (end of string)	**000001100001** **000001100010** **000001100011** **000001100100** **000001100101** **000001100110** **000000000000**
0125 0126 0127	LENGTH: .data 6,-9,15	integer 6 integer -9 integer 15	**000000000110** **111111110111** **000000001111**
0128	K: .data 22	integer 22	**000000010110**

האסמבלר מחזיק טבלה שבה רשומים כל שמות הפעולה של ההוראות והקודים הבינאריים המתאימים להם, ולכן שמות הפעולות ניתנים להמרה לבינארי בקלות. כאשר נקרא שם פעולה, אפשר פשוט לעיין בטבלה ולמצוא את הקוד הבינארי השקול.

כדי לעשות המרה לבינארי של אופרנדים שהם מענים סמליים (תוויות), יש צורך לבנות טבלה דומה. אולם בהבדל מהקודים של הפעולות, הידועים מראש, הרי המענים בזיכרון עבור הסמלים שבשימוש התכנית אינו ידוע, עד אשר התוכנית כולה נקראת ועוברת טיפול על ידי האסמבלר.

למשל, בדוגמא לעיל, האסמבלר אינו יכול לדעת שהסמל END משויך למען 117 (עשרוני), אלא רק לאחר שהתוכנית נקראה כולה.

לכן מפרידים את הטיפול של האסמבלר בסמלים לשני שלבים. בשלב הראשון בונים טבלה של כל הסמלים והערכים המספריים המשויכים להם, ובשלב השני מחליפים את כל הסמלים, המופיעים באופרנדים של הוראות התוכנית, בערכיהם המספריים. הביצוע של שני שלבים אלה כרוך בשתי סריקות (הנקראות "מעברים") של קובץ המקור.

במעבר הראשון נבנית טבלת סמלים בזיכרון, ובה לכל סמל שבתכנית המקור משוייך ערך שהוא מען בזיכרון. בדוגמא דלעיל, טבלת הסמלים לאחר מעבר ראשון היא:

סמל	ערך (בבסיס דצימלי)
MAIN	100
LOOP	103
L1	113
END	117
STR	118
LENGTH	125
K	128

במעבר השני נעשית ההמרה, כדי לתרגם את קוד המקור לקוד מכונה. בתחילת המעבר השני צריכים הערכים של הסמלים להיות כבר ידועים.

לתשומת לב: האסמבלר, על שני המעברים שלו, נועד כדי לתרגם קובץ מקור לקוד בשפת מכונה. בגמר פעולת האסמבלר, התכנית טרם מוכנה לטעינה לזיכרון לצורך ביצוע. לאחר השלמת תהליך התרגום, קוד המכונה יכול לעבור לשלבי הקישור/טעינה ולאחר מכן לשלב הביצוע.

המעבר הראשון

במעבר הראשון נדרשים כללים כדי לקבוע איזה מען ישויך לכל סמל. העיקרון הבסיסי הוא לספור את המקומות בזיכרון, אותם תופסות ההוראות. אם כל הוראה תיטען בזיכרון למקום העוקב להוראה הקודמת, תציין ספירה כזאת את מען ההוראה הבאה. הספירה נעשית על ידי האסמבלר ומוחזקת במונה ההוראות (IC). ערכו ההתחלתי של IC הוא 0, ולכן נטענת ההוראה הראשונה במען 0. ה-IC מתעדכן בכל שורת הוראה המקצה מקום בזיכרון. לאחר שהאסמבלר

קובע מהו אורך ההוראה, ה-IC מוגדל במספר התאים (מילים) הנתפסים על ידי ההוראה, וכך הוא מצביע על התא הפנוי הבא.

כאמור, כדי לקודד את ההוראות בשפת מכונה, מחזיק האסמבלר טבלה, שיש בה קוד מתאים לכל שם פעולה. בזמן התרגום מחליף האסמבלר כל שם פעולה בקוד שלה, וכל אופרנד בקידוד מתאים. אך פעולת ההחלפה אינה כה פשוטה. ההוראות משתמשות בשיטות מיעון מגוונות לאופרנדים. אותה פעולה יכולה לקבל משמעויות שונות, בכל אחת משיטות המיעון, ולכן יתאימו לה קידודים שונים לפי שיטות המיעון. לדוגמא, פעולת ההזזה mov יכולה להתייחס להעתקת תוכן תא זיכרון לאוגר, או להעתקת תוכן אוגר לאוגר אחר, וכן הלאה. לכל אפשרות כזאת של mov עשוי להתאים קידוד שונה.

על האסמבלר לסרוק את שורת ההוראה בשלמותה, ולהחליט לגבי הקידוד לפי האופרנדים. בדרך כלל מתחלק הקידוד לשדה של שם הפעולה, ושדות נוספים המכילים מידע לגבי שיטות המיעון.

במחשב שלנו קיימת גמישות לגבי שיטת המיעון של כל אחד מהאופרנדים בנפרד. <u>הערה</u>: דבר זה לא מחייב לגבי כל מחשב. ישנם מחשבים בהם, למשל, כל הפקודות הן בעלות אופרנד יחיד (והפעולות מתבצעות על אופרנד זה ועל אוגר קבוע). יש גם מחשבים עם פקודות של שלשה אופרנדים (כאשר האופרנד השלישי משמש לאחסון תוצאת הפעולה), ועוד אפשרויות אחרות.

כאשר נתקל האסמבלר בתווית המופיעה בתחילת השורה, הוא יודע שלפניו הגדרה של תווית, ואז הוא משייך לה מען – תוכנו הנוכחי של ה-IC. כך מקבלות כל התוויות את מעניהן בעת ההגדרה. תוויות אלה מוכנסות לטבלת הסמלים, המכילה בנוסף לשם התווית גם את המען ומאפיינים נוספים. כאשר תהיה התייחסות לתווית באופרנד של הוראה כלשהי, יוכל האסמבלר לשלוף את המען המתאים מטבלת הסמלים.

הוראה יכולה להתייחס גם לסמל שטרם הוגדר עד כה בתכנית, אלא יוגדר רק בהמשך התכנית. להלן לדוגמא, הוראת הסתעפות למען שמוגדר על ידי התווית A שמופיעה רק בהמשך הקוד:

```
        bne     A
        .
        .
        .
A:      ......
```

כאשר מגיע האסמבלר לשורת ההסתעפות (bne A), הוא טרם נתקל בהגדרת התווית A וכמובן לא נתן לה מען, ולכן אינו יכול להחליף את הסמל A (האופרנד של ההוראה bne) במענו בזיכרון. נראה בהמשך כיצד נפתרת בעיה זו.

בכל מקרה, תמיד אפשר לייצר במעבר הראשון את הקוד הבינארי המלא של המילה הראשונה של כל פקודה, וכן את הקוד הבינארי של כל הנתונים (המתקבלים מההנחיות data., string.).

המעבר השני

במעבר הראשון, האסמבלר אינו יכול לבנות את קוד המכונה של אופרנדים המשתמשים בסמלים שלא הוגדרו עדיין. רק לאחר שהאסמבלר עבר על כל התכנית, כך שכל התוויות נכנסו כבר לטבלת הסמלים, יכול האסמבלר להשלים את קוד המכונה של כל האופרנדים. לשם כך עובר האסמבלר שנית על כל התוכנית, ומחליף את התוויות, המופיעות באופרנדים, במעניהן המתאימים מתוך הטבלה. זהו המעבר השני, ובסופו תהיה התוכנית מתורגמת בשלמותה.

הפרדת הוראות ונתונים

בתכנית מבחינים בשני סוגים של תוכן: הוראות ונתונים. יש לארגן את קוד המכונה כך שתהיה הפרדה בין הנתונים וההוראות. הפרדת ההוראות והנתונים לקטעים שונים בזיכרון היא שיטה עדיפה על פני הצמדה של הגדרות הנתונים להוראות המשתמשות בהן.

אחת הסכנות הטמונות באי הפרדת ההוראות מהנתונים היא, שלפעמים עלול המעבד, בעקבות שגיאה לוגית בתכנית, לנסות "לבצע" את הנתונים כאילו היו הוראות חוקיות. למשל, שגיאה שיכולה לגרום תופעה כזו הסתעפות לא נכונה. התכנית כמובן לא תעבוד נכון, אך לרוב הנזק הוא יותר חמור, כי נוצרת חריגת חומרה ברגע שהמעבד מבצע פעולה שאינה חוקית.

האסמבלר שלנו חייב להפריד, בקוד המכונה שהוא מייצר, בין קטע הנתונים לקטע ההוראות. **כלומר בקובץ הפלט (בקוד המכונה) תהיה הפרדה של הוראות ונתונים לשני קטעים נפרדים, ואילו בקובץ הקלט אין חובה שתהיה הפרדה כזו.** בהמשך מתואר אלגוריתם של האסמבלר, ובו פרטים כיצד לבצע את ההפרדה.

גילוי שגיאות בתכנית המקור

האסמבלר אמור לגלות ולדווח על שגיאות בתחביר של תכנית המקור, כגון פעולה שאינה קיימת, מספר אופרנדים שגוי, סוג אופרנד שאינו מתאים לפעולה, שם אוגר לא קיים, ועוד שגיאות אחרות. כמו כן מוודא האסמבלר שכל סמל מוגדר פעם אחת בדיוק.

מכאן, שכל שגיאה המתגלה על ידי האסמבלר נגרמת (בדרך כלל) על ידי שורת קלט מסוימת.

לדוגמא, אם מופיעים שני אופרנדים בהוראה שאמור להיות בה רק אופרנד יחיד, האסמבלר ייתן הודעת שגיאה בנוסח "יותר מדי אופרנדים".

על כל הודעת שגיאה לציין גם את מספר השורה בתכנית המקור בה זוהתה השגיאה.

לתשומת לב: האסמבלר אינו עוצר את פעולתו אחרי שנמצאה השגיאה הראשונה, אלא ממשיך לעבור על הקלט כדי לגלות שגיאות נוספות, ככל שישנן. כמובן שאין כל צורך לייצר את קבצי הפלט אם נתגלו שגיאות (ממילא אי אפשר להשלים את קוד המכונה).

הטבלה הבאה מפרטת מהן של שיטות המיעון החוקיות, עבור אופרנד המקור ואופרנד היעד של ההוראות השונות הקיימות בשפה הנתונה:

שם פעולה	שיטות מיעון חוקיות עבור אופרנד מקור	שיטות מיעון חוקיות עבור אופרנד יעד
mov	1,3,5	3,5
cmp	1,3,5	1,3,5
add	1,3,5	3,5
sub	1,3,5	3,5
not	אין אופרנד מקור	3,5
clr	אין אופרנד מקור	3,5
lea	3	3,5
inc	אין אופרנד מקור	3,5
dec	אין אופרנד מקור	3,5
jmp	אין אופרנד מקור	3,5
bne	אין אופרנד מקור	3,5
red	אין אופרנד מקור	3,5
prn	אין אופרנד מקור	1,3,5
jsr	אין אופרנד מקור	3,5
rts	אין אופרנד מקור	אין אופרנד יעד
stop	אין אופרנד מקור	אין אופרנד יעד

אלגוריתם שלדי של האסמבלר

לחידוד ההבנה של תהליך העבודה של האסמבלר, נציג להלן אלגוריתם שלדי למעבר הראשון ולמעבר השני. לתשומת לב: אין חובה להשתמש דווקא באלגוריתם זה.

אנו מחלקים את קוד המכונה לשני אזורים, אזור ההוראות (code) ואזור הנתונים (data) .
לכל אזור יש מונה משלו, ונסמנם IC (מונה ההוראות - Instruction-Counter)
ו-DC (מונה הנתונים - Data-Counter).

כמו כן, נסמן ב- L את מספר המילים שתופס קוד המכונה של הוראה נתונה.

בכל מעבר מתחילים לקרוא את קובץ המקור מהתחלה.

מעבר ראשון

1. אתחל DC ← 0, IC ← 0.
2. קרא את השורה הבאה מקובץ המקור. אם נגמר קובץ המקור, עבור ל-16.
3. האם השדה הראשון הוא סמל? אם לא, עבור ל-5.
4. הדלק דגל "יש הגדרת סמל".
5. האם זוהי הנחיה לאחסון נתונים, כלומר, האם הנחית data. או string.? אם לא, עבור ל-8.
6. אם יש הגדרת סמל (תווית), הכנס אותו לטבלת הסמלים עם סימון (סמל מסוג data). ערכו יהיה DC. (אם הסמל כבר נמצא בטבלה יש להודיע על שגיאה).
7. זהה את סוג הנתונים, קודד אותם בזיכרון, עדכן את מונה הנתונים DC בהתאם לאורכם, חזור ל-2.
8. האם זו הנחית extern. או הנחית entry. ? אם לא, עבור ל-11.
9. האם זוהי הנחית extern. ? אם כן, הכנס כל סמל (אחד או יותר) המופיע כאופרנד של ההנחיה לתוך טבלת הסמלים ללא ערך, עם סימון (סמל מסוג external).
10. חזור ל-2.
11. אם יש הגדרת סמל, הכנס אותו לטבלת הסמלים עם סימון (סמל מסוג code). ערכו יהיה IC (אם הסמל כבר נמצא בטבלה יש להודיע על שגיאה).
12. חפש את שם הפעולה בטבלת שמות הפעולות, ואם לא נמצא – הודע על שגיאה בשם ההוראה.
13. נתח את מבנה האופרנדים של ההוראה וחשב את L. בנה כעת את הקוד הבינארי של המילה הראשונה של הפקודה.
14. עדכן IC ← L + IC
15. חזור ל-2.
16. אם נמצאו שגיאות בקובץ המקור, עצור.
17. עדכן בטבלת הסמלים את ערכם של הסמלים מסוג data , ע"י הוספת הערך הסופי של IC (ראו הסבר בהמשך).
18. התחל מעבר שני.

מעבר שני

1. אתחל IC ← 0
2. קרא את השורה הבאה מקובץ המקור. אם נגמר קובץ המקור, עבור ל- 10.
3. אם השדה הראשון הוא סמל, דלג עליו.
4. האם זוהי הנחיה extern. ,string. ,data.? אם כן, חזור ל-2.
5. האם זוהי הנחיה entry.? אם לא, עבור ל-7.
6. סמן בטבלת הסמלים את הסמלים המתאימים כ-entry. חזור ל-2.
7. השלם את קידוד האופרנדים החל מהמילה השניה בקוד הבינארי של ההוראה, בהתאם לשיטת המיעון. אם אופרנד הוא סמל, מצא את המען בטבלת הסמלים.
8. עדכן IC ← IC + L
9. חזור ל-2.
10. אם נמצאו שגיאות במעבר שני, עצור.
11. צור ושמור את קבצי הפלט: קובץ קוד המכונה קובץ סמלים חיצוניים, וקובץ סמלים של נקודות כניסה (ראו פרטים נוספים בהמשך).

נפעיל אלגוריתם זה על תוכנית הדוגמא שהצגנו קודם:

```
MAIN:     mov    @r3 ,LENGTH
LOOP:     jmp    L1
          prn    -5
          bne    LOOP
          sub    @r1, @r4
          bne    END
L1:       inc    K
          bne    LOOP
END:      stop
STR:      .string "abcdef"
LENGTH:   .data  6,-9,15
K:        .data  22
```

נבצע עתה מעבר ראשון על הקוד הנתון. נבנה את טבלת הסמלים. כמו כן, נבצע במעבר זה גם את קידוד כל הנתונים, וקידוד המילה הראשונה של כל הוראה. את החלקים שעדיין לא מתורגמים במעבר זה, נשאיר כמות שהם.

אנו מניחים שהקוד ייטען לזיכרון החל מהמען 100 (בבסיס דצימאלי).

Decimal Address	Source Code	Explanation	Binary Machine Code
0100 0101 0102	MAIN: mov @r3 ,LENGTH	first word of instruction source register 3 address of label LENGTH	**101000001100** **000110000000** **?**
0103 0104	LOOP: jmp L1	 address of label L1	**000100101100** **?**
0105 0106	prn -5	 immediate value -5	**000110000100** **111111101100**
0107 0108	bne LOOP	 address of label LOOP	**000101001100** **?**
0109 0110	sub @r1, @r4	 registers r1 and r4	**101001110100** **000010010000**
0111 0112	bne END	 address of label END	**000101001100** **?**
0113 0114	L1: inc K	 address of label K	**000011101100** **?**
0115 0116	bne LOOP	 address of label LOOP	**000101001100** **?**
0117	END: stop		**000111100000**
0118 0119 0120 0121 0122 0123 0124	STR: .string "abcdef"	ascii code 'a' ascii code 'b' ascii code 'c' ascii code 'd' ascii code 'e' ascii code 'f' ascii code '\0' (end of string)	**000001100001** **000001100010** **000001100011** **000001100100** **000001100101** **000001100110** **000000000000**
0125 0126 0127	LENGTH: .data 6,-9,15	integer 6 integer -9 integer 15	**000000000110** **111111110111** **000000001111**
0128	K: .data 22	integer 22	**000000010110**

טבלת הסמלים:

סמל	ערך (בבסיס דצימלי)
MAIN	100
LOOP	103
L1	113
END	117
STR	118
LENGTH	125
K	128

נבצע עתה את המעבר השני ונרשום את הקוד בצורתו הסופית:

Decimal Address	Source Code	Binary Machine Code
0100 0101 0102	MAIN: mov @r3 ,LENGTH	**101000001100** **000110000000** **000111110110**
0103 0104	LOOP: jmp L1	**000100101100** **000111000110**
0105 0106	prn -5	**000110000100** **111111101100**
0107 0108	bne LOOP	**000101001100** **000110011110**
0109 0110	sub @r1, @r4	**101001110100** **000010010000**
0111 0112	bne END	**000101001100** **000111010110**
0113 0114	L1: inc K	**000011101100** **001000000010**
0115 0116	bne LOOP	**000101001100** **000110011110**
0117	END: stop	**000111100000**
0118 0119 0120 0121 0122 0123 0124	STR: .string "abcdef"	**000001100001** **000001100010** **000001100011** **000001100100** **000001100101** **000001100110** **000000000000**
0125 0126 0127	LENGTH: .data 6,-9,15	**000000000110** **111111110111** **000000001111**
0128	K: .data 22	**000000010110**

לאחר סיום עבודת האסמבלר, קבצי הפלט מועברים לשלבי הקישור והטעינה.
לא נדון כאן באופן עבודת שלבי הקישור/טעינה (כאמור, אלה אינם למימוש בפרויקט זה).
לאחר סיום שלבים אלה, התוכנית תהיה טעונה בזיכרון ומוכנה לריצה.

קבצי קלט ופלט של האסמבלר

בהפעלה של האסמבלר, יש להעביר אליו באמצעות ארגומנטים של שורת הפקודה (command-line arguments) רשימה של שמות קבצי מקור (אחד או יותר). אלו הם קבצי טקסט, ובהם תכניות בתחביר של שפת האסמבלי, שהוגדרה למעלה. האסמבלר פועל על כל קובץ

מקור בנפרד, ויוצר עבורו קובץ מטרה (object) נפרד המכיל את קוד המכונה. כמו כן האסמבלר יוצר קובץ externals עבור כל קובץ מקור בו יש הצהרות על סמלים חיצוניים, וכן קובץ entries עבור כל קובץ מקור בו יש הצהרות על סמלים כנקודות כניסה.

שמות קבצי המקור חייבים להיות בעלי הסיומת "as.". למשל, השמות x.as , y.as, ו-hello.as הם שמות חוקיים. העברת שמות הקבצים הללו כארגומנטים לאסמבלר נעשית <u>ללא ציון הסיומת</u>.
לדוגמא: נניח שתוכנית האסמבלר שלנו נקראת assembler, אזי שורת הפקודה הבאה:

assembler x y hello

תגרום לכך שהאסמבלר יפעל על הקבצים: x.as, y.as, hello.as.

האסמבלר יוצר שמות לקבצי הפלט משם קובץ הקלט, כפי שהופיע בשורת הפקודה, בתוספת הסיומת "ob." עבור קובץ ה-object, הסיומת "ent." עבור קובץ ה-entries, והסיומת "ext." עבור קובץ ה-externals.

<u>מבנה כל קובץ פלט יתואר בהמשך</u>.

לדוגמא: הפקודה: assembler x
תיצור את הקובץ x.ob וכן את הקבצים x.ext ו-x.ent אם קיימות הצהרות של entries/externalsבקובץ המקור.

אופן פעולת האסמבלר

נרחיב כאן על אופן פעולת האסמבלר, בנוסף לאלגוריתם השלדי שניתן לעיל.

האסמבלר מחזיק שני מערכים, שייקראו להלן מערך ההוראות ומערך הנתונים. מערכים אלו נותנים למעשה תמונה של זיכרון המכונה (גודל כל כניסה במערך זהה לגודלה של מילת מכונה: 12 סיביות). במערך ההוראות מכניס האסמבלר את הקידוד של הוראות המכונה שנקראו במהלך המעבר על קובץ המקור. במערך הנתונים מכניס האסמבלר את קידוד הנתונים שנקראו מקובץ המקור (שורות מסוג data. string.).

לאסמבלר יש שני מונים: מונה ההוראות (IC) ומונה הנתונים (DC). מונים אלו מצביעים על המקום הבא הפנוי במערכים לעיל, בהתאמה. כשמתחיל האסמבלר לעבור על קובץ מקור, שני מונים אלו מאופסים.

בנוסף יש לאסמבלר טבלה, אשר בה נאספות כל התוויות בהן נתקל האסמבלר במהלך המעבר על הקובץ. לטבלה זו קוראים טבלת סמלים (symbol-table). לכל סמל (תווית) נשמרים שמו, ערכו וטיפוסו (external או relocatable).

האסמבלר קורא את קובץ המקור שורה אחר שורה, מחליט מהו סוג השורה (הערה, הוראה, הנחיה או שורה ריקה) ופועל בהתאם.

1. שורה ריקה או שורת הערה: האסמבלר מתעלם מהשורה ועובר לשורה הבאה.

2. שורת הוראה:

האסמבלר מוצא מהי הפעולה, ומהן שיטות המיעון של האופרנדים. מספר האופרנדים אותם הוא מחפש נקבע בהתאם להוראה שנמצאה. האסמבלר קובע לכל אופרנד את ערכו באופן הבא:

- אם זה אוגר – האופרנד הוא מספר האוגר.
- אם זו תווית (מיעון ישיר) – האופרנד הוא ערך התווית כפי שמופיע בטבלת הסמלים (ייתכן והסמל טרם נמצא בטבלת הסמלים).
- אם זה מספר (מיעון מיידי) – האופרנד הוא המספר עצמו.
- אם זו שיטת מיעון אחרת – ערכו של האופרנד נקבע לפי המפרט של שיטת המיעון (ראו תאור שיטות המיעון לעיל)

קביעת שיטת המיעון נעשית בהתאם לתחביר של האופרנד, כפי שהוסבר לעיל בהגדרת שיטות המיעון. למשל, מספר מציין מיעון מיידי, תווית מציינת מיעון ישיר, שם של אוגר עם @ לפניו מציין מיעון אוגר, וכד'.

לאחר שהאסמבלר ניתח את השורה והחליט לגבי הפעולה, שיטת מיעון אופרנד המקור (אם יש), ושיטת מיעון אופרנד היעד (אם יש), הוא פועל באופן הבא:

אם זוהי פעולה בעלת שני אופרנדים, אזי האסמבלר מכניס למערך ההוראות, במקום עליו מצביע מונה ההוראות IC, את קוד המילה הראשונה של ההוראה (בשיטת הייצוג של הוראות המכונה כפי שתואר קודם לכן). מילה זו מכילה את קוד הפעולה, ואת שיטות המיעון. בנוסף "משרייין" האסמבלר מקום במערך עבור המילים הנוספות הנדרשות עבור הוראה זו, ומגדיל את מונה ההוראות בהתאם. אם אחד או שני האופרנדים הם בשיטת מיעון אוגר או מיידי, האסמבלר יקודד כעת גם את המילים הנוספות הרלוונטיות במערך ההוראות.

אם זוהי פעולה בעלת אופרנד אחד בלבד, כלומר אין אופרנד מקור, אזי הקידוד הינו דומה לעיל, פרט לסיביות של שיטת המיעון של אופרנד המקור במילה הראשונה, אשר יכילו תמיד 0, מכיוון שאינן רלוונטיות לפעולה.

אם זוהי פעולה ללא אופרנדים (rts, stop) אזי תקודד רק המילה הראשונה (והיחידה). הסיביות של שיטות המיעון של שני האופרנדים יכילו 0.

אם בשורת ההוראה קיימת תווית, אזי התווית מוכנסת אל טבלת הסמלים תחת השם המתאים, ערך התווית הוא ערך מונה ההוראות לפני קידוד ההוראה, וסוג התווית הוא relocatable.

3. שורת הנחיה:

כאשר האסמבלר נתקל בהנחיה, הוא פועל בהתאם לסוג ההנחיה, באופן הבא:

I. '.data'
האסמבלר קורא את רשימת המספרים, המופיעה לאחר '.data', מכניס כל מספר אל מערך הנתונים, ומקדם את מצביע הנתונים DC באחד עבור כל מספר שהוכנס.

אם בשורה '.data' יש תווית, אזי תווית זו מוכנסת לטבלת הסמלים. היא מקבלת את הערך של מונה הנתונים DC שלפני הכנסת המספרים למערך הנתונים. הטיפוס של התווית הוא relocatable, וכמו כן מסומן שההגדרה ניתנה בחלק הנתונים.

בסוף המעבר הראשון, ערך התווית יעודכן בטבלת הסמלים על ידי הוספת ה-IC (כלומר הוספת האורך הכולל של קידוד כל ההוראות). הסיבה לכך היא שבתמונת קוד המכונה, הנתונים מופרדים מההוראות, וכל הנתונים יופיעו אחרי כל ההוראות (ראה תאור קבצי הפלט בהמשך).

II. '.string'
הטיפול ב-'.string' דומה ל- '.data', אלא שקודי ה-ascii של התווים הם אלו המוכנסים אל מערך הנתונים (כל תו בכניסה נפרדת). לאחר מכן מוכנס הערך 0 (המציין סוף מחרוזת) אל מערך הנתונים. מונה הנתונים מקודם באורך המחרוזת + 1 (כי גם האפס בסוף המחרוזת תופס מקום).

הטיפול בתווית המופיעה בשורה, זהה לטיפול הנעשה בהנחיה '.data'.

III. '.entry'
זוהי בקשה לאסמבלר להכניס את התווית המופיעה כאופרנד של '.entry' אל קובץ ה-entries. האסמבלר רושם את הבקשה ובסיום העבודה, התווית הנ"ל תירשם בקובץ ה-entries.

IV. '.extern'
זוהי הצהרה על סמל (תווית) המוגדר בקובץ אחר, ואשר קטע האסמבלי בקובץ הנוכחי עושה בו שימוש. האסמבלר מכניס את הסמל אל טבלת הסמלים. ערכו הוא 0 (הערך האמיתי לא ידוע, וייקבע רק בשלב הקישור), וטיפוסו הוא external. לא ידוע באיזה קובץ נמצאת הגדרת הסמל (וגם אין זה משנה עבור האסמבלר).

יש לשים לב: בהוראה או בהנחיה אפשר להשתמש בשם של סמל אשר ההצהרה עליו ניתנת בהמשך הקובץ (אם באופן ישיר על ידי תווית ואם באופן עקיף על ידי הנחיית extern).

פורמט קובץ ה-object

האסמבלר בונה את תמונת זיכרון המכונה כך שקידוד ההוראה הראשונה מקובץ האסמבלי **יכנס למען 100 (בבסיס 10)** בזיכרון, קידוד ההוראה השניה יכנס למען העוקב אחרי ההוראה הראשונה (תלוי במספר המילים של ההוראה הראשונה), וכך הלאה עד להוראה האחרונה.

מיד לאחר קידוד ההוראה האחרונה, מכניסים לתמונת הזיכרון את קידוד הנתונים שנבנו על ידי ההנחיות 'string.' 'data.'. הנתונים יוכנסו בסדר בו הם מופיעים בקובץ המקור. אופרנד של הוראה שמתייחס לסמל שהוגדר באותו קובץ, יקודד כך שיצביע על המקום המתאים בתמונת הזיכרון שבונה האסמבלר.

נשים לב שהמשתנים מופיעים בתמונת הזיכרון אחרי ההוראות. זוהי הסיבה בגללה יש לעדכן בטבלת הסמלים, בסוף המעבר הראשון, את ערכי הסמלים המגדירים נתונים (סמלים מסוג data.).

עקרונית, קובץ object מכיל את תמונת הזיכרון שתוארה כאן. קובץ object מורכב משורות של טקסט כדלקמן.

השורה הראשונה היא כותרת המכילה שני מספרים בבסיס 10: האורך הכולל של קטע ההוראות (במילות זיכרון), ואחריו האורך הכולל של קטע הנתונים (במילות זיכרון). בין שני המספרים יש רווח אחד

השורות הבאות מכילות את תוכן הזיכרון. בכל שורה מופע תוכן של מילה אחת, לפי הסדר החל מהמילה בכתובת 100. תוכן המילה מקודד בשיטת Base64.

ניתן לקרוא על שיטת Base64 בקישור הבא: https://en.wikipedia.org/wiki/Base64

מילה היא בגודל 12 ביטים. כל 6 ביטים יומרו לתו מתאים לפי הטבלה המגדירה את Base64. לפיכך בכל שורה בקובץ object (מלבד השורה הראשונה) יהיו בדיוק שני תווים בקוד Base64.

קובץ object לדוגמה, כפי שאמור להיבנות על ידי האסמבלר, נמצא בהמשך.

פורמט קובץ ה-entries

קובץ ה-entries בנוי משורות טקסט. כל שורה מכילה שם של סמל שהוגדר כ- entry ואת ערכו, כפי שנמצא בטבלת הסמלים. הערכים מיוצגים בבסיס 10.

פורמט קובץ externals

קובץ ה-externals בנוי אף הוא משורות טקסט. כל שורה מכילה שם של סמל שהוגדר external, וכתובת בקוד המכונה בה יש קידוד של אופרנד המתייחס לסמל זה. כמובן שייתכן ויש מספר כתובות בקוד המכונה בהם מתייחסים לאותו סמל חיצוני. לכל התייחסות כזו תהיה שורה נפרדת בקובץ ה-externals. הכתובות מיוצגות בבסיס 10

להלן קובץ מקור (קלט) לדוגמה, בשם ps.as

```
; file ps.as

.entry LENGTH
.extern W
MAIN:       mov     @r3 ,LENGTH
LOOP:       jmp     L1
            prn     -5
            bne     W
            sub     @r1, @r4
            bne     L3
L1:         inc     K
.entry LOOP
            jmp     W
END:        stop
STR:        .string "abcdef"
LENGTH:     .data   6,-9,15
K:          .data   22
.extern L3
```

להלן טבלת הקידוד המלא הבינארי שמתקבל מהפעלת האסמבלר על קובץ המקור, ולאחריה הפורמטים של קבצי הפלט השונים הנבנים על ידי האסמבלר.

Decimal Address	Source Code	Binary Machine Code
0100 0101 0102	MAIN: mov @r3 ,LENGTH	**101000001100** **000110000000** **000111110110**
0103 0104	LOOP: jmp L1	**000100101100** **000111000110**
0105 0106	prn -5	**000110000100** **111111101100**
0107 0108	bne W	**000101001100** **000000000001**
0109 0110	sub @r1, @r4	**101001110100** **000010010000**
0111 0112	bne L3	**000101001100** **000000000001**
0113 0114	L1: inc K	**000011101100** **001000000010**
0115 0116	jmp W	**000100101100** **000000000001**
0117	END: stop	**000111100000**
0118 0119 0120 0121 0122 0123 0124	STR: .string "abcdef"	**000001100001** **000001100010** **000001100011** **000001100100** **000001100101** **000001100110** **000000000000**
0125 0126 0127	LENGTH: .data 6,-9,15	**000000000110** **111111110111** **000000001111**
0128	K: .data 22	**000000010110**

הקובץ ps.ob:

נבחין כי בדוגמה לעיל, קטע ההוראות בגודל 18 מילים, ואילו קטע הנתונים בגודל 11 מילים.
להלן קובץ הפלט המתאים לדוגמה זו, מקודד בשיטת Base64 (סה"כ 30 שורות, כולל הכותרת):

```
18 11
oM
GA
H2
Es
HG
GE
/s
FM
AB
p0
CQ
FM
AB
Ds
IC
Es
AB
Hg
Bh
Bi
Bj
Bk
Bl
Bm
AA
AG
/3
AP
AW
```

הקובץ ps.ent:

כל ערכי הסמלים מיוצגים בבסיס 10.

LOOP	103
LENGTH	125

הקובץ ps.ext:

כל הכתובות מיוצגות בבסיס 10.

W	108
L3	112
W	116

לתשומת לב: אם בקובץ המקור אין הצהרת extern. אזי לא ייווצר עבורו קובץ ext. בדומה, אם אין בקובץ המקור הצהרת entry., לא ייווצר קובץ ent. אין ליצור קובץ ext או ent שנשאר ריק.

הערה: אין חשיבות לסדר השורות בקבצים מסוג ent או ext. כל שורה עומדת בפני עצמה.

סיכום והנחיות כלליות

- אורך התוכנית, הניתנת כקלט לאסמבלר, אינו ידוע מראש (ואינו קשור לגודל 1024 של הזיכרון במעבד הדמיוני). ולכן אורכה של התוכנית המתורגמת אינו אמור להיות צפוי מראש. אולם בכדי להקל במימוש התכנית, ניתן להניח גודל מקסימלי. לפיכך יש אפשרות להשתמש במערכים לאכסון קוד המכונה בלבד. מבני נתונים אחרים בפרויקט (למשל טבלת הסמלים), יש לממשל על פי יעילות/תכונות נדרשות (למשל באמצעות הקצאת זיכרון דינמי ורשימה מקושרת).

- קבצי הפלט של האסמבלר צריכים להיות בפורמטים המופיעים למעלה. שמם של קבצי הפלט צריך להיות תואם לשם קובץ הקלט, פרט לסיומות. למשל, אם קובץ הקלט הוא prog.as אזי קבצי הפלט שייווצרו הם : prog.ob, prog.ext, prog.ent .

- אופן הפעלת האסמבלר צריך להיות תואם לנדרש בממ״ן, ללא שינויים כלשהם. אין להוסיף תפריטים אינטראקטיביים למיניהם, וכד׳. הפעלת התוכנית תיעשה רק עם ארגומנטים של שורת פקודה.

- יש להקפיד לחלק את מימוש האסמבלר למספר מודולים (קבצים בשפת C) לפי משימות. אין לרכז משימות מסוגים שונים במודול יחיד. מומלץ לחלק למודולים כגון : מבני נתונים, מעבר ראשון, מעבר שני, פונקציות עזר (למשל, תרגום לבסיס, ניתוח תחבירי של שורה), טבלת סמלים, ועוד.

- יש להקפיד ולתעד את המימוש באופן מלא וברור, באמצעות הערות מפורטות בקוד.

- יש להקפיד להתעלם מרווחים וטאבים, ולהיות סלחנים כלפי תוכנית קלט העושה שימוש ביותר רווחים מהנדרש. למשל, אם בהוראה יש שני אופרנדים המופרדים בפסיק, אזי לפני או אחרי הפסיק יכולים להיות רווחים וטאבים בכל כמות. בדומה, גם לפני ואחרי שם ההוראה. בכל המקרים האלה השורה צריכה להיחשב חוקית (לפחות מבחינת הרווחים).

- במקרה שתוכנית הקלט בשפת אסמבלי מכילה שגיאות תחביריות, נדרש לגלות ולדווח, כמו באסמבלר אמיתי, על כל השגיאות הקיימות, ולא לעצור לאחר גילוי השגיאה הראשונה. כמובן שאם הקלט מכיל שגיאות, אין טעם להפיק קבצי פלט (ob, ext, ent).

תם ונשלם פרק ההסברים והגדרת הפרויקט.

בשאלות ניתן לפנות לקבוצת הדיון באתר הקורס, ואל כל אחד מהמנחים בשעות הקבלה שלהם.

להזכירכם, באפשרותו של כל סטודנט לפנות לכל מנחה, לאו דווקא למנחה הקבוצה שלו, לקבלת עזרה. שוב מומלץ לכל אלה שטרם בדקו את התכנים באתר הקורס לעשות זאת. נשאלות באתר זה הרבה שאלות בנושא חומר הלימוד והממ״נים, והתשובות יכולות להועיל לכולם.

לתשומת לבכם : לא תינתן דחיה בהגשת הממ״ן, פרט למקרים מיוחדים (כגון מילואים או מחלה). במקרים כאלה יש לבקש ולקבל אישור מראש מצוות הקורס.

ב ה צ ל ח ה !