

RSS: תקן להפצה ושיתוף תכנים באינטרנט

24/04/2003

משה הלוי

halemo@actcom.co.il

המושג RSS הוא מושג חדש יחסית בזירת האינטרנט. בגדול, המושג RSS מתאר דרך להפצת תכנים של אתר אינטרנט כלשהו, בדרך שתהיה מובנת לא רק למשתמש, אלא גם למכונות ולמחשבים. בניגוד לתכנים באינטרנט הנמצאים על דף במבנה HTML, הניתן לצפייה בעזרת תוכנת דפדפן ביתית המותקנת על מחשב, תכנים המופצים במבנה RSS, הם תכנים שגם מכונות יכולות להבין, ולהסיק מהם מה היא הכותרת, מה הוא התיאור של התוכן, מי כתב אותו, מתי הוא נכתב והיכן הוא נמצא על הרשת.

הקדמה

לפני פריצתו של הדפדפן הציבורי הראשון, דפדפן בשם מוזאיק (mosaic), ואחריו דפדפן נטסקייפ (netscape), השירותים הפופולריים ביותר להפצת תכנים ושיתופם בין משתמשי האינטרנט שעדיין לא נקראו "גולשים", היו תוכנות הדואר האלקטרוני ותוכנות שאפשרו קריאה בקבוצות דיון ברשת.

תוכנת הדואר האלקטרוני אפשרה משלוח טקסטים בין שני משתמשים או בין משתמש אחד ליותר משתמשים. משתמש אינטרנט שלא היה בקבוצה, לא יכול היה לקרוא את התכנים שנכתבו, אלא אם כן ביקש ממישהו לשלוח לו (forward) לכתובת הדואר האלקטרונית שלו.

היסטוריה של הפצת תכנים ברשת האינטרנט

על מנת להפיץ תכנים לכולם, פותח שירות דומה לדואר האלקטרוני בשם יוזנט (usenet), ראשי תיבות של "רשת משתמשים" (user's network). יוזנט הכילה קבוצות דיון (newsgroups) רבות בנושאים שונים ומגוונים וכל משתמש יכול היה להתחבר ולקרוא דיונים קודמים שבהם לא היה שותף, וכן לשלוח את תגובותיו ותכניו לאותה קבוצה. משלוח התוכן אל אותה קבוצה היה דומה בעיקרו למשלוח דואר אלקטרוני, רק שבמקום כתובת דואר אלקטרוני של המקבל, נרשמה כתובת קבוצת הדיון.

למשל, אם משתמש כלשהו רצה להשתתף בקבוצת דיון בנושא "העברית של תוכנת חלונות 98 של מיקרוסופט", הוא היה בוחר את הקבוצה הבאה:
microsoft.public.il.hebrew.win98

קבוצת דיון זו שכנה תחת כתובת שרת מחשב כלשהו ברשת, בדרך כלל שרת קבוצות דיון של ספק שירות האינטרנט (ISP) או האוניברסיטה (לפני שרשת האינטרנט נפתחה לציבור הרחב). כך למשל, אם ספק השירות היתה חברה בשם "מיקרוסופט", כתובת השרת שהכיל את רשימת קבוצות הדיון שכן תחת הכתובת news.microsoft.com.

בהשלכה לימינו אנו, ימים בהם לכל דבר ברשת יש כתובת URL הרי כתובת אינטרנט זו נראית כך:

<news://news.microsoft.com/microsoft.public.il.hebrew.win98>

דרך אחרת, פחות פופולרית, להצגת הודעה בודדת מתוך קבוצות הדיון היא:
[nntp://news.microsoft.com/microsoft.public.il.hebrew.win98/article2334](http://news.microsoft.com/microsoft.public.il.hebrew.win98/article2334)

במקום הקידומת http המציינת שמדובר בפרוטוקול תקשורת (של משפחת tcp/ip) להעברת טקסט עם קישורי על (hyper text transport protocol), ציין פרוטוקול news או בשמו הרשמי nntp ראשי תיבות של Network News Transfer Protocol.

על מנת להשתתף ולקרוא תכנים, היה על המשתמש להתקין תוכנה (תוכנת לקוח), דומה לתוכנה לקריאת דואר אלקטרוני, שבהפעלתה התחברה אל שרת מחשב מרוחק וקראה ממנו את הכותרות. כאשר המשתמש בחר בכותרת מסוימת, התוכנה פנתה שוב אל השרת המרוחק, והביאה את התוכן במלואו אל מחשב המשתמש. כך יכול היה המשתמש לקרוא תכנים שהופצו ברשת לרבים.

כיום, תוכנות לקריאת קבוצות דיון של יוזנט, משולבות ביחד עם תוכנות לקריאת דואר אלקטרוני כמו אאוטלוק אקספרס (Outlook Express) של מיקרוסופט. הדמיון הטכני הרב בין צורת הקריאה של דואר אלקטרוני לבין קריאת תוכן של קבוצת דיון, מאפשרות שילוב של שני שירותי האינטרנט הללו בתוכנה אחת.

מה שקרה בהמשך לקבוצות הדיון של יוזנט, הוא מה שקרה לדואר האלקטרוני עצמו. בחודש פברואר 2001, קנתה חברת גוגל, בעלת מנוע החיפוש הטוב ביותר כיום ברשת, את ארכיב קבוצות הדיון של חברת דג'ה ניוז (Deja News), ובכך נתנה גם למשתמשי האינטרנט, גולשי הדפדפנים, את האפשרות לצפות ולהשתתף בעשרות אלפי קבוצות דיון ללא צורך בהתקנת תוכנה מיוחדת על המחשב שלהם.

בחודש מאי שנת 2001, גוגל שמה על הרשת את כל קבוצות הדיון בהן יותר מ 650 מיליון הודעות, וכל זאת עם אפשרות לצפות ולעיין בהודעות בעשרת תוכנת דפדפן פשוטה. הבונבון בכל המהלך הזה, היתה האפשרות לחפש תכנים שהיו קיימים על שרתי קבוצות דיון ברשת כבר בשנות ה 80, שנים רבות לפני קיומם של אתרי האינטרנט.

כיום, כשאתם מחפשים מידע כלשהו במנוע החיפוש של גוגל, תוכלו למצוא את אותו מידע גם בקבוצות הדיון. המידע בקבוצות הדיון הוא לפעמים הרבה יותר מפורט, ענייני, לא עטוף בגרפיקה מפריעה ובפרסומות, ונכתב ע"י אנשים שאכפת להם לתת תשובות.

פריצת הדפדפן לאוויר העולם

ביום 22 אפריל 1993 פרץ לעולם הדפדפן הציבורי הראשון, דפדפן בשם מוזאיק (Mosaic). דפדפנים פנימיים ופשוטים כבר היו ברשות אנשי האקדמיה, בהם שיתפו מסמכים שכתבו (אז הם לא נקראו דפי אינטרנט), וניתן היה לגשת אליהם ולקרוא את אותם מסמכים שנכתבו כמאמרים המכילים קישורי על (הייפר טקסט), טקסטים בהם חלק מהמילים מקושרות למסמכים אחרים שכתבו אחרים (מראי מקום).

את דפדפן מוזאיק פיתחה קבוצת סטודנטים באוניברסיטת אילינוי (University of Illinois). המיוחד בדפדפן הזה היתה האפשרות שלו לעבוד תחת מערכת הפעלה חלונאית עם שימוש בעכבר, כמו windows גרסה 3 שהיתה נפוצה באותה תקופה בקרב הציבור הרחב,

פרוטוקול HTTP שנמצא בשימוש בדפדפן ובשימוש שרתי המחשב המארחים דפים, היה קיים כבר בשנת 1990, ונכתב ע"י קבוצה שתמכה ברעיון של "מארג רשת עולמי" (World Wide Web). רק בחודש מאי 1996 פורסם רשמית מסמך המתאר את פרוטוקול התקשורת

(RFC1945), וכך דפי האינטרנט, בנוסף לדפדפן, זכו להכרה בינלאומית של הגופים המנהלים את רשת האינטרנט.

את רעיון רשת ה WWW הגה טים ברנרס לי (Tim Berners Lee) כבר בשנת 1989. המצאת הדפדפן אפשרה לשתף מאמרים בין עמיתים בצורה נוחה לקריאה ולעבודה עימם (ע"י לחציה על קישורים ודפדוף בין מסמכים).

הדפדפן וה WWW הפכו ליישום המחץ (killer application) של השימוש באינטרנט, כשם שבשנות ה 80 המוקדמות הפכה תוכנת הגיליון האלקטרוני ליישום המחץ וגרמה לעליית המכירות של מחשבים אישיים. הדפדפנים ואתרי האינטרנט הפכו לכל כך נפוצים וחזקים, עד שהיה צורך להמציא את הגלגל מחדש, על מנת לא להשתמש בכלים ישנים לקריאה ולשיתוף תוכן כשם שבוצע בעבר ע"י תוכנות לקריאת קבוצות דיון.

דף עם מידע על טים ברנרס לי, ממציא ה WWW
<http://www.w3.org/People/Berners-Lee>

כך למשל, קמו להם פורומים וקבוצות דיון ברשת האינטרנט המבוססים על דפי אינטרנט המשתנים באופן דינמי בהתאם להודעות הנשלחות. בהמשך, קמו להם גם רשימות דיוור וקבוצות, אליהם כל משתמש אינטרנט שנקרא כעת גולש, יכול היה להירשם ולקבל בתיבת הדואר האלקטרונית שלו את כל המכתבים שכתבו אחרים. המשתמש היה יכול להגיב למכתב, ואת תגובתו יכולו לראות כולם, גם מי שלא נרשם לקבוצה, אם הקבוצה נוהלה באתר אינטרנט, כמו בפורטל yahoo למשל.

אז מה רע במסמכים ובתוכן שמופץ בדפי אינטרנט ?

למרות הפופולריות שלו זכה הדפדפן וצורת העבודה שלו בהצגת מידע על רשת האינטרנט, המסמכים שהופצו ברשת האינטרנט באתרי האינטרנט, לא היו "מוגדרים היטב" כמו המסמכים שהופצו בקבוצות הדיון. כך למשל, מסמך הקיים בדף אינטרנט, אינו מגדיר בתוכו את הכותרת האמיתית שלו, את שם המחבר, את תאריך ההפקה ואת תיאור המסמך. כך למשל, בדף כלשהו, כל הכותרות יודגשו, כאשר בדף אינטרנט אחר, תחת הכותרות יושם קו תחתון.

למרות זאת, יש דפי אינטרנט שכן מכילים מידע זה, תחת תגים שנקראים תגי meta הנמצאים בראשית המסמך, ויכולים להגדיר את שם המחבר, שם המסמך, תאריך הפקת המסמך, השפה שבה כתוב המסמך ועוד פרמטרים שונים.

דוגמה לתגי meta בדף במבנה HTML:

```
<META NAME="Subject" CONTENT="Games">
<META NAME="Date" CONTENT="20 Nov 2000">
<META NAME="Keywords" CONTENT="Tetris, Palman, Space Invaders">
<META NAME="Description" CONTENT="Web site of Games">
<META NAME="Copyright" CONTENT="Koko Games Inc">
<META NAME="Author" CONTENT="Sami Saviv">
<META NAME="Revisit-after" CONTENT="30">
<META NAME="Webmaster" CONTENT="Steven Spilberg">
```

מכונות ותוכנות על מחשב יכולות לקרוא דף כזה ולנתח את הנתונים מתוך הדף. הבעיה היחידה היא שאין תקן מוסכם על משמעות התגים. כך למשל, יכול להיות טקסט המתאר את תאריך המסמך מסומן ע"י המאפיין "date" או "pubDate" או אפילו "TheDate".

כותרת המסמך יכולה להיות מסומנת כ "title" או כ "subject" או סתם "headline". זאת בניגוד למסמכים המופצים בדואר אלקטרוני ובקבוצות דיון, שם לעולם תהיה כותרת המסמך תחת השם "subject". כתובת שולח המסמך תמיד תהיה תחת השם "from" וכך גם לגבי תאריך המסמך שנמצא תמיד תחת שדה "date". כך יכולות תוכנות דואר כמו אאוטלוק אקספרס ואחרות, לנתח את המידע ולהציגו בטבלה מתאימה לשימוש המשתמש.

דוגמה לכותרות של דואר אלקטרוני, מוגדרות היטב:

שם כותב ההודעה וכתובתו תמיד יהיו תחת שדה From.

שם מקבל ההודעה וכתובתו תמיד יהיו תחת שדה To.

נושא ההודעה תמיד יהיה תחת שדה Subject.

תאריך ההודעה תמיד יהיה תחת שדה Date ובפורמט ידוע ומוגדר.

```
From: "Koko" <koko@koko.net>
To: "halemo" <halemo@actcom.co.il>
Subject: need your advise
Date: Wed, 23 Apr 2003 09:46:02 +0200
MIME-Version: 1.0
```

Hi,

What age should people start sex?

Koko

מכונה או מחשב המפענחים את המסמך, חייבים לדעת היטב מה הם התגים והמאפיינים המתאימים המתארים את תוכן המסמך. לפיכך, אתר תוכן הרוצה למשל שמנוע חיפוש כלשהו ינתח את הדף שבנה, חייב לקבל על עצמו את ההגדרות הפרטיות של מנוע החיפוש. כך למשל, מנוע חיפוש אחד יבקש לשים את שם המחבר תחת המאפיין "Author", ומנוע חיפוש אחר יבקש לשים את שם המחבר תחת המאפיין "Writer". אין תקן ברור ומוגדר היטב לתג ולמאפיינים.

המצאתה של שפת XML

כדי לפתור את הבעיה של ניתוח המידע הקיים בדפי אינטרנט שהפכו לנפוצים ביותר ברשת, פותחה שפת סימון תכנים בשם XML, ראשי תיבות של eXtensible Markup Language, כלומר "שפת סימון ברת הרחבה".

בשנת 1996 פיתחה קבוצת חברים בארגון W3C את שפת XML. בשנת 1998 הציגה קבוצת החברים בארגון W3C, ארגון המארגן את התקנים של רשת ה WWW, את המפרט הראשון שהגדיר שפת הסימון XML. המסמך נערך ונכתב ע"י שלושה חברים מקבוצת העבודה: טים בריי (Tim Bray), נציג חברת נטסקייפ, ג'ין פולי (Jean Paoli), נציג חברת מיקרוסופט, וסי אם ספנסר מקווין (C. M. Sperberg-McQueen) מאוניברסיטת אילינוי שבשיקגו (שם כזכור הומצא הדפדפן הציבורי הראשון בשם מוזאיק).

מפרט XML הראשון שנכתב בשנת 1998
<http://www.w3.org/TR/1998/REC-xml-19980210>

המסמך תיאר את שפת XML כשפה ממשיכה של שפת SGML שהיא תקן בינלאומי של ISO (קרי: איזו), המשמשת לתיאור מידע במסמכים (תקן ISO8879). שפת SGML, ראשי תיבות של Structure Generalized Markup Language, ובעברית "שפת סימון כללית בעלת מבנה" היא למעשה שפת האם הקדמונית של שפת HTML, וממנה נגזרת גם שפת XML. בניגוד לשפת SGML שהיא רעיון כללי, שפת HTML היא הת'כלס של הרעיון ומשמשת לבניית דפי אינטרנט, לניתוחם ולקריאתם ע"י תוכנת דפדפן.

מפרט XML תיאר את צורת ייחוס המשמעות לנתונים. כך למשל, אם בשפת HTML, טקסט כלשהו נעטף בתג שאפשר לטקסט להיות מודגש כאשר הוצג בדפדפן, כך שפת XML תיארה ייחוס משמעות לטקסט.

לדוגמה:

בשפת HTML הטקסט הבא יהיה מודגש (bold):

```
<b>Hello world</b>
```

ובשפת XML, יקבל הטקסט משמעות שהוא "שם של ספר":

```
<bookTitle>Hello world</bookTitle>
```

דוגמה למסמך XML בסיסי

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<note>
  <to>Madonna</to>
  <from>halemo</from>
  <heading>Need Your Advise</heading>
  <body>Would you like to have sex with me ?</body>
</note>
<book>
  <title>Sex for dummies</title>
  <Author>Arik Pin</Author>
  <price>100.00</price>
</book>
```

אתרי התוכן משתמשים בשפת XML

שפת XML נתנה פתרון כלשהו לאתרי תוכן שרצו להפיץ את התכנים שלהם לעולם הרחב ולשימושם של מכונות ומחשבים. כך יכול היה אתר התוכן ליצור סינדיקציה (שיתוף פעולה) של תכנים בינו לבין מקורות ששילמו על המידע ובחרו להציג אותו בדרך שונה מהדרך המקורית שהתוכן הוצג באתר התוכן היוצר.

גורם כלשהו שרצה לינוק תכנים מאתר התוכן, היה חייב לעבוד על פי ההגדרות שניתנו לו ע"י אותו אתר תוכן. כך, למעשה, שוב מתקיימת הבעיה של הגדרת משמעות המידע. באתר תוכן א, כותרת המידע תהיה תחת התג "headline", ובאתר תוכן אחר שגם הוא מפיץ את

תכניו באמצעות שפת XML, יבחר את התג "title" בה ישים את הטקסט המתאר את כותרת המאמר.

כך למעשה, אם היו עשרה אתרי תוכן שהפיצו את התוכן שלהם באמצעות שפת XML, היה צריך ליצור עשרה מפענחי תוכן שונים.

שפת XML לא הגדירה במפורש את שמות התגים שציינו את מהות המידע כלפי גורם צד שלישי, אלא הגדירה שיטה להצגת משמעות המידע והתוכן.

נטסקייפ ויוזרלנד ממציאות את RSS

בחודש יוני 1999 הציגה חברת נטסקייפ לעולם את שפת RSS גרסה 0.9. ראשי התיבות של RSS של נטסקייפ היו Rich Site Summary, שמשמעותו "תמצית אתר עשיר (בתוכן)".

הרעיון של נטסקייפ היה ליצור שיטה ומסגרת ליצירת פורטל תוכן שיאסוף תכנים מכל מיני מקורות ברשת, ינתח אותם לפי הכותרת שלהם, שם המחבר, תאריך יצירת המסך וקישור אל המסמך המקורי.

פורטל איסוף התוכן של נטסקייפ
<http://my.netscape.com>

הרעיון של נטסקייפ, הוא ה RSS בגרסתו הראשונה, גרסה 0.9, היה ליצור קובץ טקסט שישב על השרת בכל אתר אינטרנט המעונין להפיץ את תכניו ונטסקייפ תוכל לקרוא אותו ולייצר דף אינטרנט שייציג את הכותרות והתיאורים וכמובן קישור אל התוכן. מבנה הקובץ היה דומה מאוד למבנה XML, למעשה התבסס עליו, אך היה פשוט הרבה יותר במבנה שלו.

הגרסה הבאה של RSS, גרסה 0.91, שאבה את המבנה העדכני שלה מפרוייקט של חברת יוזרלנד (UserLand), שכבר בחודש דצמבר 1997, פרסמה פורמט ליצירת סינדיקציה של תכנים ברשת האינטרנט. הפורמט של חברת יוזרלנד (UserLand) נקרא ScriptingNews.

גרסה 0.91 של RSS הוסיפה להגדרות גם את תגית description שאפשרה בפועל שיתוף ממשי של תכנים, כי כעת לא רק הכותרת והקישור לקובץ היו חלק מקובץ ה RSS, אלא גם תיאור קצר של התוכן. כך, אתרי אינטרנט שאספו חדשות, יכלו להציג בנוסף לכותרת ולקישור, גם תיאור קצר של התוכן, דבר משמעותי בפני עצמו.

UserLand ScriptingNews
[http://my.userland.com/stories/storyReader\\$11](http://my.userland.com/stories/storyReader$11)

ההגדרות של התגים בקובץ RSS, הגדרות כתובות היטב, אפשרו לכל מי שרצה לאסוף תוכן או להציג תוכן, ליצור קובץ שבו לתגים משמעות אחת ויחידה. כך למשל, הטקסט של כותרת המאמר תמיד יהיה תחת התג title. הקישור אל המאמר תמיד יהיה תחת התג link (ולא למשל url או siteurl). תאריך פרסום המאמר תמיד יהיה תחת pubDate, וגם כאן, מוגדר התאריך בפורמט מסוים וברור שאותו יכולה לקרוא המכונה או תוכנת המחשב.

ההגדרות הברורות היטב של הנתונים בקובץ RSS והמשמעות לכל תג ותג, יצרו למעשה קובץ XML מוגדר היטב, שכל מי שרצה יכול היה להתחבר אליו, בלי לתאם מראש עמדות ונתונים מול יצרן התוכן.

גולשים שיצרו תכנים ופרסמו תכנים ברשת האינטרנט, באתרים ובפורומים, יכלו למעשה להפיץ את תורתם תחת מבנה מוגדר היטב, כך שכל מכונה יכולה להבין. פורמט RSS יצר מצב שיצרן התוכן לא צריך לדעת מי קורא את התוכן שלו, וקורא התוכן לא חייב לספר ליצרן שהוא או המכונה שלו קוראים את התוכן שלו.

דוגמה לדף בפורמט RSS עם עברית:

```
<?xml version="1.0" encoding="windows-1255"?>
<rss version="0.91">

<channel>
  <title>HALEMO::חדר המידע</title>
  <link>http://halemo.net/info/index.html</link>
  <description>
חדר המידע, מידע בעברית על דברים שלא קיימים ברשת בעברית, וגם
קצת מאמרים
  </description>
  <language>he</language>

  <item>
    <pubDate>2003-04-08T00:00:00+02:00</pubDate>
    <title>תתאבד</title>
    <link>http://halemo.net/info/views/0003.html</link>
    <description>
המצב הכלכלי הקשה גורם לאנשים להתאבד. המצוקה הכלכלית של האזרח
הקטן, נגרמת עקב חוסר הכנסה, רדיפה מצד השלטונות וגופים חזקים
בגין חובות, ובאין אפשרות למצוא מקור פרנסה בגלל מדיניות
? הממשלה, האדם בוחר במוות. אולי מישהו אחר צריך להתאבד
    </description>
  </item>

  <item>
    <pubDate>2003-04-03T00:00:00+02:00</pubDate>
    <title>סע בביטחון - סע אגד</title>
    <link>http://halemo.net/info/egged/index.html</link>
    <description>
בשנות ה 30 ו ה 40, בתקופת המנדט הבריטי, כשכל תושבי פלסטינה,
יהודים וערבים, היו פלשטינאים, ואוטובוסים לא היו מתפוצצים על
ידי מחבלים מתאבדים, הפיק קואופרטיב אגד כרזה פרסומית הקוראת
לציבור הנוסעים לבוא ולנצל את שירותי התחבורה של אגד. סיסמת
הקמפיין שלוותה את הכרזה המצויירת היתה: סע בבטחון, סע אגד
    </description>
  </item>

</channel>
</rss>
```

אז בשביל מה זה טוב ?

בעולם האינטרנט וגם בעולם הרגיל, קיימים אנשים רבים שכותבים ומייצרים תוכן שלא לשם רווח. לאותם אנשים חשוב שאנשים אחרים יוכלו לדעת שהם הוציאו לאור מאמר חדש.

קיימים גם גופים שכן מרוויחים מיצירה ומהפצת תכנים, אך מעוניינים לשחרר מקצת התכנים לציבור הרחב ללא תשלום, לשם יצירת נוכחות ברשת וכדי שיכירו אותם ואת כתיבתם.

בשנתיים האחרונות גדלה קרנה של קהילת הבלוגרים, גולשים אשר כותבים על עצמם ועל אירועים שמעניינים אותם ביומן אינטרנט (web log) אישי. קהילת קוראי הבלוגים גם היא גדלה, וכיום יש מצב שיש כותבי בלוגים ויש קוראי בלוגים ויש יחס של יצרן תוכן וצרכן תוכן.

כאשר יש מצב שיש עודף מידע והרבה אתרים שיש לסרוק אותם ולקרוא את תוכנם, נוצר מצב שהקורא אינו מעוניין לבזבז את זמנו בסיוור באתרי אינטרנט, והוא חייב כלי שיאסוף עבורו את המקורות הרלוונטיים וייתן לו קישורים אל התכנים הרלוונטיים. כמו כן, אם המשתמש רוצה לשמור על קישורים לתכנים מעניינים, הוא אינו חייב להשתמש בדפדפן כדי לשמור "מועדפים" או Favorites (בדפדפן אינטרנט אקספלורר) או bookmarks (בדפדפן של נטסקייפ).

למעשה, תוכנת לקריאת RSS מאפשרת גם ניהול של קישורים מעניינים. ברצותו, יכול הגולש להמשיך ולשמור וברצותו, למחוק את הקישור, בדיוק כמו בעבודה עם תוכנת דואר אלקטרוני. דבר מאפשר גם לחפש כותרות במחשב המקומי, אם התוכנה לקריאת RSS מאפשרת זאת (זה אפשרי מבחינה טכנית). המשתמש אינו צריך לחפש את המידע שקרא קודם לכן במנועי חיפוש ברשת כמו גוגל ודומיו.

פורמט ה-RSS פתר בעיה זו, כאשר תוכנות לקריאת קבצים בפורמט RSS, שהוא פורמט XML מוגדר היטב, החלו לצוץ ברחבי הרשת. התוכנות לקריאת קבצי RSS, קבצים הנמצאים באתרי אינטרנט, הן תוכנות הדומות בצורתן לתוכנות לקריאת דואר אלקטרוני כמו אאוטלוק אקספרס של מיקרוסופט. מצד שמאל בעמוד התוכנה קיים עץ עם שמות נושאים ושמות תיקיות. בצד ימין, הגדול יותר בחלקו היחסי על המסך, קיימת למעלה רשימת הנושאים הכוללת לפחות כותרת ותאריך המאמר, ובתחתיתה קיים חלון דפדפן המציג את הדף הרלוונטי כאשר לוחצים על הכותרת. בהקלקה פעמיים עם העכבר, יכול התוכן להיפתח בדפדפן רגיל.

קבצי RSS הנמצאים באתר האינטרנט שמפיץ את התוכן (ולרוב גם מייצר אותו), מתעדכנים מדי זמן. התוכנות לקריאת קבצי ה-RSS מנתחות אילו תכנים חדשים נוספו ומעדכנות בהתאם את התצוגה.

הטרמינולוגיה (אוצר המילים) של ה-RSS

שימוש ב-RSS פיתח גם טרמינולוגיה ענפה שלא השתמשת בה כדי לא להכביד על הקורא שלו זה נושא חדש. להלן המושגים השונים:

Feed

פעולה של הוספת רשומה לקובץ ה-RSS, הכוללת לפחות את כותרת המאמר והקישור אל המאמר. קובץ RSS שנמצא באתר אינטרנט, נקרא לרוב RSS Feed, מכיוון שהוא "מזין" (Feed פירושו "הזנה") תוכן אל מכונות אחרות או תוכנות לקוח שקוראות קבצי RSS. קובץ ה-RSS נקרא Feed.

Post

משלוח ידיעה, מאמר או כתבה לאתר תוכן כלשהו. כאשר משתמש קורא את אותה ידיעה או מאמר או כתבה, הוא בעצם קורא את ה"פוסט" של מחבר התוכן.

RDF

Resource Description Framework

החל מגרסה 1.0 של RSS, ראשי התיבות של המונח RSS שונו מ Rich Site Summary לראשי התיבות של RDF Site Summary ובעברית "תמצית אתר RDF". המונח RDF הוא ראשי תיבות של Resource Description Framework, ובעברית "מסגרת לתיאור משאבים". כך הפך לו RSS של גרסאות 0.9x למונח חדש ששמו הוא "תמצית אתר המתאר מסגרת לתיאור משאבים".

RSS

ראשי תיבות של:

Rich Site Summary (ר"ת של RSS גרסאות 0.9x)

RDF Site Summary (ר"ת של RSS בגרסה 1.0)

Really Simple Syndication (ר"ת של RSS גרסה 2.0)

המונח השלישי (גרסה 2.0), מתאר ביטוי שמשמעו "סינדיקציה מאוד פשוטה (של תכנים)"

Channel

ערוץ תוכן. בערוץ, שמוגדר בקובץ ה RSS תחת תג בשם channel, מוגדרים שם הערוץ (title), כתובת הערוץ באינטרנט (link), תיאור הערוץ (description), השפה בה כתובים תכני הערוץ (language). כמו כן מוגדרים כל התכנים שכרגע נמצאים בערוץ, תחת התג item. בפועל, "ערוץ" הוא למעשה "דף או אתר אינטרנט".

Item

פריט מידע. תחת תג item מפורטים התכנים המופצים. כל פריט מכיל כותרת (title) וקישור (link). נתונים נוספים שיכולים להיכלל הם תאריך היצירה (pubDate), שם המחבר (Author) ותיאור המאמר (description).

RSS Reader

תוכנה שתפקידה לקרוא ולנתח את קבצי ה RSS ולהציג את התכנים החדשים.

RSS Client

כמו RSS Reader. תוכנת לקוח שנמצאת אצל המשתמש שקורא את קבצי ה RSS.

RSS Server

המחשב שמספק את התכנים בפורמט RSS. למעשה, זה אתר האינטרנט שעליו נמצא קובץ ה RSS, והקישור אליו ניתן לתוכנה שתפקידה לקרוא ולנתח את ה RSS.

האם יש ל RSS סיכוי ?

רעיון ה RSS עדיין בחיתוליו במונחי זמן של התפתחויות טכנולוגיות ברשת האינטרנט. כבר היום ניתן לראות את הצלחתם של אתרי אינטרנט שאוספים כותרות חדשותיות של אתרים שונים ומציגים אותם לקורא בצורה נוחה ועניינית.

כך למשל, התפתחו להם הפורומים באינטרנט. הפורומים הם למעשה גרסה וואבית של קבוצות הדיון הישנות שעבדו תחת יוזנט. הפורומים המצליחים ביותר הם פורומים אשר מציגים למשתמש רשימה של כותרות עם תאריכים. המשתמש, כיום נקרא "הגולש", לוחץ על הכותרת המעניינת אותו, וההודעה נפתחת וניתן לקרוא את התוכן שמתחת לכותרת וגם את התגובות של הקוראים האחרים.

למעשה, כל אתר בוחר להציג את הכותרת ואת התאריך בדרכו שלו בהתאם לתוכנה שאותה הוא מריץ אצלו בשרת שבאתר האינטרנט שהוא מחזיק. החסרון הוא בכך שרק התוכנה של האתר יכולה לדעת ולהבין היכן הכותרת והיכן התאריך. כל אדם אחר גם כן יכול להבין זאת. אבל, מכונה חיצונית או תוכנה חיצונית אינה יכולה להבין היכן נכתב התאריך והיכן נכתבה הכותרת. כאשר מידע נוסף מצורף לשורת המידע, הדבר מסתבך יותר.

כל אתר אינטרנט וכל מערכת פורומים בוחרים להציג מידע באופן שונה, למרות שכביכול מדובר באותו מידע. בקובץ בפורמט RSS זה לא יכול לקרות. מיקום התאריך ומיקום הכותרת מוגדרים היטב. מבנה התאריך ידוע והוא ניתן לפענוח ע"י מכונה או ע"י תוכנה כלשהי.

דוגמה:

פורום העניין שלנו באתר "הייד פארק"
<http://news.hydepark.co.il>

דוגמה אחרת לדף אינטרנט מצליח שמיישם (בלי לדעת) את הרעיון מאחורי RSS הוא דף מבזקים, שאוסף כותרות מאתרי אינטרנט שונים ומציג אותם ברשימה מסודרת הכוללת את הכותרת, את התאריך, את הקישור לידיעה ונתונים נוספים כגון האתר (המחבר) של הידיעה. בדפים אילו אוספת תוכנה שנכתבה במיוחד כדי לפענח ולאסוף ידיעות מאתרי אינטרנט שונים. כאשר אתר חדשות כלשהו ישנה את הפורמט שלו, עלול להיווצר מצב שהידיעה לא תופיע במבזק, כי התוכנה לא הצליחה לפענח היכן הכותרת והיכן התאריך.

דוגמה:

המבזקים של פרש
<http://news.fresh.co.il>

דוגמה לדף עם מבנה של מבזקים

מדענים עיראקים קיבלו הוראה להשמיד סיגריות		10:25:57	23/04/03
חיייל התהפך עם אופניים צבאיות		0:12:53	23/04/03
ילד נשך כלב שנבח על הזתול שלו		10:10:11	22/04/03
אלמונים שדדו את הליצן של מקדונלדס		10:09:00	21/04/03
ראש הממשלה יתפטר בשל המצב הכלכלי		08:03:22	21/04/03

היתרונות והרעיונות המרכזיים של השימוש ב RSS

אוטומציה של תכנים

מתן אפשרות למכונות ולתוכנות מחשב לקרוא תמצית של תכנים ולהציגם למשתמש בכל דרך שנראית לתוכנה או למכונה, מבלי שתהיה התערבות כלשהי בעיצוב התוכן. כך למשל, יכולה מכונה או תוכנת מחשב להציג את התאריך משמאל לכותרת. תוכנה אחרת, תומכת עברית למשל, תוכל להציג את התאריך מימין לכותרת התוכן.

מכונות להצגה ולטיפול בתוכן יכולות להיות מדפסות חכמות, מכשירי כף יד (פאלם פיילוט), פקסימיליה חכמה וכדומה.

שימוש ישיר ברשת האינטרנט וחסכון בעלויות

למפיץ התכנים אין צורך בהקמת שרת נוסף לשם הפצה ושיתוף בתכנים. בניגוד לעבר שהקמת קבוצות דיון הצריכה שימוש בתוכנה מיוחדת וברשת חדשות עם תמיכה בפרוטוקול תקשורת nntp ובתמיכה ביציאת תקשורת (port) מספר 119, בהפצת תכנים דרך RSS, נעשה בשירות שהוא נפוץ וקיים כבר ואינו מצריך עליות נוספות.

קובץ ה-RSS הוא למעשה קובץ שנמצא באתר אינטרנט. הפנייה אל אתר האינטרנט נעשית דרך דפדפן, אבל יכולה להיעשות דרך תוכנה אחרת שתבצע תקשורת בפרוטוקול http, בשימוש ביציאת תקשורת (port) מספר 80 שפתוחה כל הזמן בשרת האינטרנט.

ביטול התלות בין יוצר התוכן לקורא התוכן

בעל אתר אינטרנט המבקש להפיץ את התוכן שלו דרך אתר האינטרנט שלו, אינו חייב ליידע את ספק האינטרנט שלו או את זה שנותן לו את שירות האחסון. קובץ ה-RSS הוא למעשה דף אינטרנט שנמצא באתר אינטרנט ונותר לספק רק קישור אליו. מפיץ התוכן מפרסם את הקישור לקובץ ה-RSS לכל, וכל מי שחפץ להירשם, אינו חייב ליידע את מפיץ התכנים. הוא פשוט מעתיק את הקישור לתוכנה לקריאת RSS והתוכנה כבר מטפלת בכותרות. מכיוון שהתגים בקובץ RSS מגדירות היטב את מהות הטקסטים, אין צורך לברר כל תג מהו אצל מפיץ התוכן. מדובר בשיתוף פעולה אנונימי.

קלות השימוש

שפת RSS היא קלה יחסית לכתיבה. המשתמש אינו צריך לגמור אוניברסיטה במדעי המחשב, כדי לקרוא דף המגדיר את התכנים המופצים.

קלות היצירה

ניתן לכתוב את המידע בקובץ RSS באמצעות עורך טקסטים רגיל או מעבד תמלילים המפיק טקסט נקי או באמצעות תוכנה כלשהי המאפשרת לכתוב מידע. גם סקריפטים (PHP, ASP, JSP, Perl) הנמצאים בשרתי אינטרנט, יכולים להפיק בצורה אוטומטית, קבצים בפורמט RSS. אתרי תוכן שעובדים עם מערכת תוכן מבוססת בסיס נתונים, יכולים גם הם להפיק קובץ כזה משדות שכבר מוגדרים במערכת ניהול התוכן. מכיוון שקל מאוד לכתוב ידנית את הקובץ, גם אתרי תוכן סטטיים שאינם תומכים בסקריפטים, יכולים לשים באתר קובץ RSS שכזה ולהפיץ את התכנים שלהם.

קלות עיבוד ע"י תוכנה או מכונה

התוכנה המשמשת לקריאת התכנים, אינה צריכה להיות מורכבת ומסובכת. ניתוח קובץ במבנה RSS, הוא קל יחסית, ואינו מצריך עליות גבוהות או משאבי זיכרון וקוד תוכנה מסובך.

קלות קריאה ע"י אדם

מסמכי RSS הם קבצים טקסטואליים (ולא קבצים בינאריים). הם קלים לקריאה על ידי כל אדם בצורה סבירה. קובץ RSS מאפשר למשתמש לקרוא את הכותרות, התאריכים ותיאור המסמך, גם ללא תוכנת מיוחדת, אלא בעזרת דפדפן או תוכנה לעריכת טקסטים.

ניהול תוכן קל ופשוט

מסמכי RSS הם מסמכים המגדירים היטב מה הכותרת, מה תאריך היצירה, שם הכותב, קישור אל המאמר ברשת ומידע נוסף. הגדרות אליו מאפשרות ניהול אישי של תוכן שאחרים כתבו, וגם ניהול של תוכן אישי שאדם כתב והפיץ. מכיוון שפורמט RSS הוא פורמט מוגדר היטב ונפוץ, ניתן להתחבר (ליצור ממשק) אל מערכות אחרות המפצות תכנים או מחפשות תכנים.

החסרונות בשימוש ב RSS

הפצת תכנים חד צדדית

הקורא אינו יכול להגיב ישירות על תוכן המסמך, אלא יכול לצרף את תגובתו תחת הכתבה אם יש מערכת תגובות לאתר שבו מאוכסן התוכן.

חוסר משוב

מחבר או מפיץ התוכן אינו יכול תמיד לדעת מי קרא את התוכן שלו. בניגוד לדואר אלקטרוני או לקבוצות דיון של יוזנט, הקורא אינו יכול להגיב דרך RSS, אלא רק דרך משלוח דואר אלקטרוני לכותב או מפיץ התוכן או באמצעות תגובה (טוקבק) אם קיימת כזו מתחת לכתבה. בשל מצב זה, יבחרו משתמשים רבים לקרוא תוכן בלי לתת משוב לתוכן. כמו כן, בעל האתר לא ידע מהיכן נכנסו אל האתר שלו, אלא אם כן התוכנה לקריאת התוכן תספק מידע כזה (והיא לא חייבת לעשות דבר כזה). דפדפן בסיסי מספק מידע לאתר תוכן מהיכן נכנסו אליו דרך תקשורת HTTP באמצעות משתנה בשם Referrer.

חוסר תמיכה בפרסומות

אתרי תוכן המספקים את התכנים שלהם בחינם ברשת האינטרנט, מעוניינים להמשיך ולספק את התכנים בפורמט שלהם, בשילוב פרסומות מהן הם מרוויחים כסף. הפצת התכנים ללא פרסומות מסביב למאמר, תגדע את מטה לחמם, ולא תספק להם מוטיבציה להפיץ תכנים (בחינם) באמצעות RSS.

כתובות של אתרי תוכן ישראלים עם תמיכה ב RSS

בישראל קיימים מספר אתרי תוכן המספקים קבצי RSS לשימושם של תוכנות לקריאת RSS. להלן רשימה שידועה לי בזמן כתיבת שורות אלו. המדובר באתרים המפיצים קבצי RSS בגרסאות 0.91 ושאר גרסאות x0.9.

חדר המידע

<http://halemo.net/info/rss.xml>

דואר חשמלי

<http://halemo.net/edoar/rss.xml>

הגדה השמאלית

<http://www.hagada.org.il/hagada/html/backend.php>

מגזין מחשבים חודשי PCPLUS מבית אנשים ומחשבים

<http://www.pcplus.co.il/rss.php>

חדשות קונספציה

http://www.exego.net/news/news_rss.asp

חנן כהן, מידע דיגיטלי. בלוג על נושא ה RSS בפורמט RSS

<http://www.info.org.il/rss/rss.php>

תוכנות לקריאת קבצי RSS

קיימות מספר תוכנות לקריאת קבצי RSS. לחלקן בעיות ילדות כמו חוסר תמיכה בעברית כנונה כמו יישור מימין לשמאל (RTL), או חוסר תמיכה בעברית שנכתבת בתקן יוניקוד (UTF8).

בחלק מהתוכנות שבחנתי יש באגים כמו שיתוק מערכת ההפעלה, אי שחרור זכרון וסתם חוסר יכולת להתמודד עם קבצי RSS שאינם כתובים היטב.

שתי תוכנות מומלצות הן:

NewsCrawler

אחת התוכנות הטובות שנתקלתי בהן. כוללת כבר הגדרות של ערוצי תוכן כמו ה BBC ואחרים. אינה תומכת בעברית בתקן יוניקוד (16 ביט), אבל תומכת בעברית רגילה (ASCII 8 ביט)

<http://www.newzcrawler.com/downloads.shtml>

FeedReader

תוכנה פשוטה קטנה ויעילה.

<http://www.feedreader.com/downloads.php>

פיתוח RSS בגרסאות מתקדמות יותר: גרסה 1.0

כמו כל דבר טוב, קל ופשוט שמצליח, גם בנושא ה RSS יש התקדמות לעבר המורכבות. אם בעבר, פורמט HTML היה קל ופשוט, וכל מתכנת פרטי היה יכול לבנות דפדפן שקורא ומנתח קבצי HTML ומציג אותם על המסך, הרי התפתחותו של פורמט ה HTML והוספת תמיכה באובייקטים, קבצי פלאש, עיצוב CSS, תמיכה בקוד VB (ויזואל בייסיק של מיקרוסופט), תמיכה ב JAVA, תמיכה ב JavaScript ועוד, יצרה בעצם מצב שרק חברות גדולות כמו מיקרוסופט, או קבוצה גדולה של מתכנתים (כמו פרוייקט מוזילה), יכולים לבנות דפדפן ברמה שיתמוך בכל המורכבויות הללו.

פורמט RSS התפתח וכיום יש לו גרסאות מורכבות יותר ומסובכות יותר, שגם אינן תמיד שומרות על תאימות אחורה. כך למשל, פורמט RSS גרסה 1.0 של ארגון ה W3C נראה כיום כמו XML מורכב יותר. תוכנות לקריאת RSS במבנה של גרסה 1.0 חייבות להתייחס מחדש לחלק מהתגים החדשים ולנתח תגים שלא היו קיימים קודם.

מסמך הגדרת RSS גרסה 1.0

<http://web.resource.org/rss/1.0/spec>

דוגמה לדף בפורמט RSS גרסה 1.0 עם עברית:

```
<?xml version="1.0" encoding="windows-1255"?>

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"
>

<channel rdf:about=" http://halemo.net/info/about.html">
  <title>HALEMO:::חידע החידע</title>
  <link>http://halemo.net/info/index.html</link>
```

```

<description>
חדר המידע, מידע בעברית על דברים שלא קיימים ברשת בעברית, וגם
קצת מאמרים
</description>
<language>he</language>
</channel>

<item rdf:about="http://">
  <pubDate>2003-04-08T00:00:00+02:00</pubDate>
  <title>אריק תתאבד</title>
  <link>http://halemo.net/info/views/0003.html</link>
  <description>
המצב הכלכלי הקשה גורם לאנשים להתאבד. המצוקה הכלכלית של האזרח
הקטן, נגרמת עקב חוסר הכנסה, רדיפה מצד השלטונות וגופים חזקים
בגין חובות, ובאין אפשרות למצוא מקור פרנסה בגלל מדיניות
? הממשלה, האדם בוחר במוות. אולי מישהו אחר צריך להתאבד
  </description>
</item>

<item rdf:about="http://">
  <pubDate>2003-04-03T00:00:00+02:00</pubDate>
  <title>סע בביתחון - סע אגד</title>
  <link>http://halemo.net/info/egged/index.html</link>
  <description>
בשנות ה 30 ו ה 40, בתקופת המנדט הבריטי, כשכל תושבי פלסטינה,
יהודים וערבים, היו פלשטינאים, ואוטובוסים לא היו מתפוצצים על
ידי מחבלים מתאבדים, הפיק קואופרטיב אגד כרזה פרסומית הקוראת
לציבור הנוסעים לבוא ולנצל את שירותי התחבורה של אגד. סיממת
הקמפיין שלוותה את הכרזה המצויירת היתה: סע בבטחון, סע אגד
  </description>
</item>
</rdf:RDF>

```

פורמט RSS גרסה 1.0 הוסיף תמיכה בתווים בתקן יוניקוד UTF8, כך שעברית הכתובת ביוניקוד (תו של 16 ביט) יכולה היות מוצגת רק בגרסאות 1.0 ומעלה, ולא בגרסאות 0.9 שאינן תומכות ביוניקוד (UNICODE), אלא רק בתווים בגודל 8 ביט (ASCII).

פיתוח RSS בגרסאות מתקדמות יותר: גרסה 2.0

חברת יוזרלנד (UserLand), שהיתה הראשונה בשנת 1997 להגדיר תקן להפצת תכנים ברשת האינטרנט באמצעות קובץ בדומה בפורמט שלו ל XML, ראתה כי טוב וכי ה RSS קונה לו מעריצים ומשתמשים, ופיתחה את גרסה 2.0 של תקן RSS. גרסה זו אינה שייכת ל W3C שמגדיר ומרכז עבודה על פיתוח תקנים לרשת ה WWW.

בתקן 2.0 הוסיפה חברת יוזרלנד תגים נוספים לשימושם של תוכנות לקריאת קבצי RSS ואיסוף חדשות. בין התגים החדשים: בעל זכויות היוצרים (copyright), שם וכתובת מנהל התוכן (managingEditor), שם הוובמסטר וכתובתו (webMaster), קטגוריה (category) ושם התוכנה שייצרה את הקובץ (generator). קיימים תגים נוספים שאינם חייבים להיות מוצגים. התאימות אחורה לפורמט RSS בגרסאות 0.9 נשמרת.

מסמך הגדרת RSS גרסה 2.0
<http://backend.userland.com/rss>

