Word2Vec – שיטה של “word embedding” – ייצוג של מילים/ביטויים באמצעות וקטור של מספרים ממשיים. מטרת השיטה היא לייצג את המילים בצורה בעלת משמעות תחבירית/סמנטית מה שלא קיים בשיטה של “one hot”. השיטה מסתמכת על תאוריה בבלשנות לפיה מילים המופיעות לעיתים תכופות באותו הקשר הן בעלות משמעות סמנטית משותפת. בפועל הכניסה לרשת היא וקטורי "one hot" של המילים, ישנה שכבה אחת של מטריצה בגודל של מספר המילים במילון x מימד הווקטור לייצוג, שהיא בעצם lookup table עבור מילות הכניסה, כך שכל שורה היא הווקטור המתאים למילה, ובסוף פונקציית מחיר.

לשיטה שתי גרסאות: CBOW ו- .Skip-Gram הראשונה מנסה ללמוד מתוך חלון מסביב למילה את המילה, והשנייה מנסה מתוך המילה למצוא את מילות ההקשר בחלון. מנסים להביא למקסימום את פונקציית ההסתברות של מילות ההקשר מתוך המילה, עבור כל המילים בטקסט. (ובאופן שקול את פונקציית ה-“log-likelihood”). את ההסתברות מגדירים על ידי פונקציית softmax על המכפלה הסקלרית בין הווקטורים (שמסמנת את הדמיון ביניהם). משתמשים באלגוריתם אופטימיזציה (למשל SGD) כדי למזער את פונקציית המחיר (negative log likelihood). כדי לשפר אופטימיזציה משתמשים בשני ייצוגים לכל מילה – פעם אחת בתור מילת הקשר ופעם בתור מילת מטרה, ולבסוף ממצעים ביניהן.

חישוב הגרדיאנט של ה-softmax יקר ולכן משתמשים בשיטה של “Negative Sampling” כך שבמקום לחשב גרדיאנט לפי כל המילים במילון מנסים להבחין בין מילות הקשר אמתיות לבין מילות "רעש" רנדומליות, זאת באמצעות רגרסיה לוגיסטית .

כמו כן בשביל להימנע ממילים שחוזרות על עצמם הרבה בלי לתרום מידע (למשל “the”) נעשית דגימה לפי התכיפות של המילה בכל קטע הטקסט.