Demetris
Kai-
zer

Προβλέψη κατηγορίων σε άρθρα ειδήσεων

Demetris Kaizer

ΠΕΡΙΛΗΨΗ

Σοφτωαρε τεστινγ ις α προμισινγ τεχηνιχυε φορ δισχοερινγ υνκνοων υλνεραβιλιτιες ιν προγραμς. Ιν παρτιχυλαρ, σοφτωαρε τεστινγ ηας βεεν ρεαλιζεδ ιν τηε φορμ *οφ φυζζινγ* οφ νατιε χοδε, ωηερε σοφτωαρε ις εξερχισεδ υσινγ α αστ αμουντ οφ ινπυτς φορ ινφερρινγ ιφ ανψ οφ τηεμ ιντροδυχες σεχυριτψ-ρελατεδ σιδε εφφεχτς. Φορ ινστανχε, α προγραμ χραση ωηεν προχεσσινγ α γιεν ινπυτ μαψ βε α σιγναλ φορ μεμορψ-χορρυπτιον υλνεραβιλιτψ.

Αλτηουγη φυζζινγ ις σιγνιφιχαντλψ εολεδ ιν αναλψζινγ νατιε χοδε, ωεβ αππλιχατιονς ηαε ρεχειεδ λιμιτεδ αττεντιον, σο φαρ. Ιν τηις παπερ, ωε δεσιγν, ιμπλεμεντ ανδ εαλυατε ωεβΦυζζ, ωηιχη ις, το τηε βεστ οφ ουρ κνοωλεδγε, τηε φιρστ *γραψ-βοξ* φυζζερ φορ ωεβ αππλιχατιονς. ωεβΦυζζ, βεφορε αναλψζινγ α ωεβ αππλιχατιον, χαρεφυλλψ ινστρυμεντς αλλ εξεχυτινγ χοδε ιν ορδερ το χρεατε α *φεεδβαχκ λοοπ* βετωεεν τηε φυζζερ ανδ τηε αναλψζεδ σοφτωαρε. ωεβΦυζζ συππορτς ινστρυμεντινγ ΠΗΠ ανδ ΗΑΤΚ αππλιχατιονς.

Μορεοερ, ωε προιδε τηε φιρστ αττεμπτ φορ αυτοματιχαλλψ σψντηεσιζινγ ρεφλεχτιε Ροσσ-σιτε Σχριπτινγ (ΞΣΣΣ) υλνεραβιλιτιες ιν ανιλλα ωεβ αππλιχατιονς. Υσινγ αν αδδιτιοναλ ινστρυμεντατιον πασς, ωε χαν συχχεσσφυλλψ ινθεχτ ΞΣΣΣ βυγς ιν ωεβ αππς ανδ τηεν ινστρυχτ ωεβΦυζζ το δισχοερ τηεμ. Ωε δεμονστρατε βοτη βυγ ινθεχτιον ανδ ωεβΦυζζ υσινγ Ωορδ-Πρεσς ανδ Δρυπαλ. ωεβΦυζζ χαν συχχεσσφυλλψ δισχοερ ουρ ινθεχτεδ βυγς φαστερ τηαν οτηερ βλαχκ-βοξ φυζζερς.

Α"Μ Ρεφερενχε Φορματ:
Demetris Kaizer. 2021. Προβλέψη κατηγορίων σε άρθρα ειδήσεων. Ιν *Προχεεδινγς οφ Α"Μ ὃνφερενχε (ὃνφερενχε'17).* Α"Μ, Νεω Ψορκ, ΝΨ, ΥΣΑ, ;; παγες. ηττπς://δοι.οργ/10.1145/νννννν.νννννν

1 ΙΝΤΡΟΔΥ"ΤΙΟΝ

Ιν ρεχεντ ψεαρς, φυζζινγ ηας βεχομε αν εσσεντιαλ παρτ φορ δισχοερινγ βυγς ιν σοφτωαρε. Αυτοματεδ σοφτωαρε τεστινγ ορ φυζζινγ ις τηε προχεσς οφ γενερατινγ ορ μυτατινγ ινπυτς ανδ φεεδινγ τηεμ το προγραμς φορ τηε πυρποσε οφ δισχοερινγ βυγς. Αλλ αλονγ, τηερε ωας τηε νεεδ φορ αν αυτοματεδ ωαψ το δισχοερ βυγς ιν σοφτωαρε βυτ ωηατ ρεαλλψ σπαρχεδ τηε ιντερεστ φορ φυζζινγ ωας τηε ιντροδυχτιον οφ ΑΦΛ [;], τηε στατε-οφ-τηε-αρτ φυζζερ τηατ προδυχες φεεδβαχκ δυρινγ φυζζινγ βψ λεεραγινγ ινστρυμεντατιον οφ τηε αναλψζεδ προγραμ. Βψ χρεατινγ τηις *φεεδβαχκ λοοπ*, φυζζερς χαν γρεατλψ ιμπροε τηειρ περφορμανχε ας τηεψ χαν δετερμινε ωηετηερ αν ινπυτ ις ιντερεστινγ, ναμελψ ιτ τριγγερς α νεω χοδε πατη, ανδ υσε τηατ ινπυτ το προδυχε οτηερ τεστ χασες.

Σοφτωαρε τεστινγ πλαψς α σιγνιφιχαντ ρολε ιν σοφτωαρε δεελοπμεντ χψχλε βεχαυσε ωηεν υλνεραβιλιτιες αρε πρεσεντ, τηεψ χαν ηαε σεερε χονσεχυενχες. Βψ εξπλοιτινγ σοφτωαρε βυγς, αδερσαριες χαν περφορμ δατα βρεαχηες, ινσταλλ μαλιχιους μαλωαρε ορ εεν τακε χομπλετε χοντρολ οφ α δειχε. Ηοωεερ, φινδινγ βυγς βεφορε τηεψ βεχομε εξπλοιτς ις ποσσιβλε ωηιλε αλσο βεινγ α χηαλλενγινγ τασκ. Μαινλψ βεχαυσε βυγς αρε τριγγερεδ ωηεν αν υνεξπεχτεδ ινπυτ ις γιεν το τηε προγραμ, σομετηινγ ωηιχη ις διφφιχυλτ το φυλλψ σιμυλατε τηρουγη στατιχαλλψ ωριττεν υνιτ τεστς [;].

Αδδιτιοναλλψ, ωηιλε αυτοματεδ σοφτωαρε τεστινγ ηας βεχομε αν αττραχτιε φιελδ οφ ρεσεαρχη, ιτ στιλλ ηας α λονγ ωαψ το γο, εσπεχιαλλψ φορ ωεβ αππλιχατιονς [;]. Ας τηε Ιντερνετ ινφραστρυχτυρε προγρεσσες, ιτ χαν βε νοτεδ τηατ αν ινχρεασινγ νυμβερ οφ σοφτωαρε ωριττεν ιν νατιε χοδε, μιγρατες το ωεβ αππλιχατιονς. Ας α ρεσυλτ, τηις αττραχτς μορε αττα χκερς το ταργετ ωεβ αππλιχατιονς ιν ορδερ το αχηιεε τηειρ γοαλς. Τηυς, τηερε ις α γροωινγ νεεδ φορ δεελοπμεντ οφ αυτοματεδ υλνεραβιλιτιες σχαννερς τηατ ταργετ ωεβ αππλιχατιονς ας ωελλ ας φορ αυτοματεδ υλνεραβιλιτιες ινθεχτιον τοολς το εαλυατε τηε φορμερ.

Νυμερους φυζζερς ηαε βεεν δεελοπεδ ιν τηε παστ φεω ψεαρς τηατ τρψ το οπτιμιζε τηε φυζζινγ προχεσς βψ προποσινγ αριους μετηοδολογιες [; ; ; ; ; ; ; ;]. Φορ ινστανχε, μοστ οφ τηε φυζζερς τακε αδανταγε οφ ινστρυμεντατιον ον τηε σουρχε ορ βιναρψ λεελ. Τηατ ις, ινσερτινγ χοδε το τηε προγραμ ιν ορδερ το ρεχειε φεεδβαχκ ωηεν α χοδε βλοχκ γετς τριγγερεδ ανδ τρψ το αδϑυστ τηε γενερατεδ ινπυτς το ιμπροε χοδε χοεραγε. Οτηερς υτιλιζε χονχολις/σψμβολις εξεχυτιον ιν ορδερ το εξτραχτ υσεφυλ ινφορματιον αβουτ τηε προγραμ ανδ υσε τηατ ινφορματιον φορ ιμπροινγ τηε ινπυτ γενερατιον προχεσς [; ; ;]. Ηοωεερ, αλλ τηεσε φυζζερς αρε χυρρεντλψ ταργετεδ τοωαρδς φινδινγ υλνεραβιλιτιες ιν νατιε χοδε, ωηιλε ωεβ αππλιχατιονς ηαε ρεχειεδ λιμιτεδ αττεντιον.

Ιν τηις παπερ, ωε προποσε ωεβΦυζζ, ωηιχη ις, το τηε βεστ οφ ουρ κνοωλεδγε, τηε φιρστ *γρεψ-βοξ* φυζζερ φορ ωεβ αππλιχατιονς. Ὗρρεντλψ ααιλαβλε φυζζερς φορ ωεβ αππλιχατιονς αχτ ιν α βλαχκ-βοξ φασηιον [;]· τηεψ ϑυστ βρυτε φορχε τηε ταργετ ωιτη ΥΡΛς τηατ εμβεδ κνοων ωεβ-αττα χκς παψλοαδς. Ιν χοντραστ, ωεβΦυζζ φιρστλψ ινστρυμεντς α ωεβ αππλιχατιον βψ αδδινγ χοδε τηατ τραχκς αλλ χοντρολ φλοως τριγγερεδ βψ αν ινπυτ ανδ νοτιφιες τηε φυζζερ, αχχορδινγλψ. Νοτιφιχατιονς χαν βε εμβεδδεδ ιν τηε ωεβ αππλιχατιον'ς ΗΤΤΠ ρεσπονσε υσινγ χυστομ ηεαδερς ορ χαν βε ουτπυττεδ το α σηαρεδ φιλε ορ μεμορψ ρεγιον. Ον τηε οτηερ ηανδ, τηε φυζζερ σταρτς σενδινγ ρεχυεστς το τηε ταργετ ανδ αναλψζες τηε ρεσπονσες ιν ορδερ το ρεαλιζε ανψ ιντερεστινγ ρεχυεστς τηατ ωουλδ λατερ ηελπ το ιμπροε τηε χοδε χοεραγε ανδ ας α ρεσυλτ, τριγγερ υλνεραβιλιτιες νεστεδ δεεπ ιν τηε ωεβ αππλιχατιον'ς χοδε.

ὃνφερενχε'17, Θυλψ 2017, Ωασηινγτον, Δ", ΥΣΑ

Α"Μ ΙΣΒΝ 978-ξ-ξξξξ-ξξξξ-ξ/ΨΨ/ΜΜ. . . $15.00
ηττπς://δοι.οργ/10.1145/νννννν.νννννν

Ινστρυμεντατιον οφ τηε αναλψζεδ προγραμ ις *κεψ* το τηε φυζζινγ προςεσς, σινςε ιτ αλλοως τηε φυζζερ το ινσταντιατε μορε εφφιςιεντ στρατεγιες φορ μυτατινγ ινπυτς ανδ τηυς εξπλορε ας μυςη ας ποσσιβλε οφ τηε ςοδε οφ τηε αππλιςατιον. Φορ νατιε αππλιςατιονς, ινστρυμεντατιον ις ςαρριεδ ουτ ατ τηε ιντερμεδιατε ρεπρεσεντατιον οφ τηε αππλιςατιον'ς ςοδε (ε.γ., ατ τηε ΛΛ"Μ'ς ΙΡ), ωηερε τηε σουρςε ςοδε ις ααιλαβλε, ορ διρεςτλψ το τηε βιναρψ [;]. Φορ ωεβ αππλιςατιονς, ινστρυμεντατιον ις ςηαλλενγινγ, σινςε (α) σεεραλ διφφερεντ φραμεωορκς αρε υσεδ το ρεαλιζε ωεβ αππλιςατιονς, (β) αππλιςατιονς αρε εξεςυτεδ τηρουγη α ωεβ σερερ ανδ (ς) τηερε ις νο στανδαρδ ιντερμεδιατε ρεπρεσεντατιον οφ ωεβ ςοδε. ωεβΦυζζ αππλιες αλλ ινστρυμεντατιον ατ τηε αβστραςτ-σψνταξ τρεε λαψερ οφ ΠΗΠ αππλιςατιονς ανδ ςαν αλσο ινστρυμεντ ατ τηε ΗΗ"Μ λαψερ αππλιςατιονς ωριττεν ιν Ηαςκ[;]. Τηερεφορε, ουρ ινστρυμεντατιον ςαν ςοερ α σιγνιφιςαντ αμουντ οφ ααιλαβλε ωεβ ςοδε, ωηιλε ιτ ις γενερις ενουγη – λαβελινγ βασις βλοςκς, ςολλεςτινγ φεεδβαςκ, ανδ εμβεδδινγ φεεδβαςκ υσινγ ΗΤΤΠ ηεαδερς ορ οτηερ σηαρεδ ρεσουρςες αρε αλλ εντιρελψ τρανσπαρεντ φεατυρες τηατ αρε ΠΗΠ/Ηαςκ αγνοστις – το βε αππλιεδ ον οτηερ σψστεμς, ασσυμινγ τηε υνδερλψινγ τοολς φορ προςεσσινγ ςοδε αρε ααιλαβλε.

Εαλυατινγ φυζζινγ ις ανοτηερ ςηαλλενγινγ τασκ [;], σινςε μιγρατινγ κνοων υλνεραβιλιτιες το εξιστινγ σοφτωαρε, ιν ορδερ το τεστ τηε ςαπαβιλιτιες οφ τηε φυζζερ ιν φινδινγ βυγς, ςαν βε α τεδιους προςεσς [;]. Τηυς, φορ εαλυατινγ ωεβΦυζζ, βυτ αλσο οτηερ φυζζερς φορ ωεβ αππλιςατιονς, ωε δεελοπ α μετηοδολογψ φορ αυτοματιςαλλψ ινθεςτινγ βυγς ιν ωεβ αππλιςατιονς ωριττεν ιν ΠΗΠ. Ουρ μετηοδολογψ ις ινσπιρεδ βψ ΛΑ"Α [;] ανδ ταργετς ωεβ αππλιςατιονς ινστεαδ οφ νατιε ςοδε. Ινθεςτινγ υλνεραβιλιτιες ιν ωεβ ςοδε, αγαιν, ις ςηαλλενγινγ, σινςε ιμπορταντ τοολς υσεδ φορ αναλψζινγ νατιε ςοδε ανδ ινθεςτινγ υλνεραβιλιτιες (ε.γ., ταιντ-τρασκινγ ανδ ινφορματιονφλοω φραμεωορκς), αρε νοτ ααιλαβλε φορ ωεβ αππλιςατιονς. Το οερςομε τηις λασκ οφ ααιλαβλε τοολς, ουρ υλνεραβιλιτψ ινθεςτιον μετηοδολογψ λεεραγες τηε ινστρυμεντατιον ινφρασστρυςτυρε ωε υσε φορ βυιλδινγ ωεβΦυζζ, ιν τηε φιρστ πλαςε.

1.1 ὂντριβυτιονς

Ιν τηις παπερ, ωε μακε τηε φολλοωινγ ςοντριβυτιονς.

(1) Ωε δεσιγν, ιμπλεμεντ ανδ εαλυατε ωεβΦυζζ, τηε φιρστ γρεψ-βοξ φυζζερ ρεαλιζεδ φορ δισςοερινγ υλνεραβιλιτιες ιν ωεβ αππλιςατιονς. ωεβΦυζζ αππλιες ινστρυμεντατιον ον τηε ταργετ ωεβ αππλιςατιον φορ γυιδινγ τηε εντιρε φυζζινγ προςεσς. Ινστρυμεντατιον ςαν βε αππλιεδ ον τηε ΑΣΤ λεελ οφ ΠΗΠ-βασεδ ορ ον τηε ΗΗ"Μ βψτεςοδε ον Ηαςκ-βασεδ ωεβ αππλιςατιονς φορ ςρεατινγ α *φεεδβαςκ λοοπ* ανδ υτιλιζινγ ιτ ιν ορδερ το ινςρεασε ςοδε ςοεραγε. Τηις φεεδβαςκ λοοπ ις εσταβλισηεδ υσινγ ΗΤΤΠ ςυστομ ηεαδερς. ὂνσεχυεντλψ, βψ λεεραγινγ τηε φεεδβαςκ λοοπ, ωεβΦυζζ ινςρεασες τηε νυμβερ οφ ποτεντιαλ υλνεραβιλιτιες τριγγερεδ.

(2) Ωε δεσιγν ανδ ιμπλεμεντ α μετηοδολογψ φορ αυτοματεδ βυγ ινθεςτιον ιν ωεβ αππλιςατιονς ωριττεν ιν ΠΗΠ. Φορ ινθεςτινγ αρτιφιςιαλλψ ςρεατεδ υλνεραβιλιτιες, ωε ςραωλ ωεβΦυζζ-βασεδ ινστρυμεντεδ αππλιςατιονς ανδ ωε ινσερτ βυγς ιν πλαςες τηατ ςαν βε ποτεντιαλλψ εξεςυτεδ. Ουρ βυγ-ινθεςτιον μετηοδολογψ ις *νοτ ονλψ* εσσεντιαλ φορ εαλυατινγ ωεβΦυζζ βυτ αλσο ιταλ φορ τηε προγρεσσιον οφ φυρτηερ ρεσεαρςη ιν υλνεραβιλιτψ φινδινγ φορ ωεβ σοφτωαρε.

(3) Ωε τηορουγηλψ εαλυατε ωεβΦυζζ ιν τερμς οφ ςοεραγε, τηρουγηπυτ ανδ εφφιςιενςψ ιν φινδινγ υνκνοων βυγς. Φορ βεττερ υνδερστανδινγ τηε μεασυρεδ ςαπαβιλιτιες οφ ωεβΦυζζ ωε ςομπαρε ουρ ρεσυλτς ωιτη τηρεε εξιστινγ ωεβ-αππλιςατιον φυζζερς. ωεβΦυζζ ις τηε ονλψ φυζζερ τηατ ρεπορτς ςοεραγε ινφορματιον· ιν παρτιςυλαρ, ωεβΦυζζ ςαν ςοερ αβουτ 21.5% οφ τηε εντιρε ΩορδΠρεσς ςοδε, ωηιςη ςονταινς αρουνδ *ηαλφ α μιλλιον* Λο"ς, ιν 50 ηουρς οφ φυζζινγ. Ας εξπεςτεδ, ωεβΦυζζ ις σλοωερ, ιν τερμς οφ τηρουγηπυτ, δυε το τηε ινολεδ ινστρυμεντατιον. Ιν φαςτ, ανοτηερ ποπυλαρ φυζζερ, ΩφυζζΔ [;] ις τηρεε τιμες φαστερ ωηεν φυζζινγ Δρυπαλ, βυτ τηις ις σομετηινγ το βε εξπεςτεδ, σινςε τηε ρεδυςτιον οφ τηε τηρουγηπυτ δυε το τηε ινστρυμεντατιον παψς οφφ ιν ινςρεασεδ ςοεραγε ιν τηε λονγ ρυν. Φιναλλψ, ωεβΦυζζ, ςομπαρεδ το τηε οτηερ τηρεε φυζζερς, φινδς τηε μοστ ινθεςτεδ υλνεραβιλιτιες (30 ωιτη τηε σεςονδ ονε βεινγ ΩφυζζΔ ωιτη 28) φορ α φυζζινγ σεσσιον τηατ λαστς 65 ηουρς.

(4) Το φοστερ φυρτηερ ρεσεαρςη ιν τηε φιελδ, ωε ρελεασε αλλ οφ ουρ ςοντριβυτιονς, ναμελψ τηε τοολςηαιν φορ ινστρυμεντινγ ΠΗΠ/Ηαςκ αππλιςατιονς, τηε αςτυαλ φυζζερ, ανδ τηε τοολςηαιν φορ ινθεςτινγ βυγς ιν ωεβ αππλιςατιονς, ας οπεν σουρςε.