Een olifant in de garage

Zet zeven man in een garagebox en laat ze tot Kerst sleutelen aan een groots project: dat is Hansken in een notendop. Daar, in de ‘bubble’, ontwerpt, onderzoekt, bouwt en leert een multidisciplinair team aan de opvolger voor Xiraf – de forensische zoekmachine die door de gigantische hoeveelheid data in haar voegen kraakt. **Door Sebastiaan van der Lubben**

Een gemiddelde zaak brengt zo'n vier terabyte (dat is vier keer 1000 gigabyte, of een vier met twaalf nullen aan bytes) data met zich mee. Dat is niet meer handmatig uit te lezen. Vroeger, weet Harm van Beek (Digitale Technologie en Biometrie), printten rechercheurs nog wel eens vijf ordners e-mails uit. "Die werden ook heel goed gelezen." Maar doorrechercheren was er eigenlijk niet bij. "Dook er een nieuwe naam op waarvan je als rechercheur ook het een en ander wilde weten, dan was je zomaar een half jaar verder", vertelt collega Erwin van Eijk (senior wetenschappelijk medewerker). Om nog maar te zwijgen over de eindeloze chatsessies, internetbezoekjes, plaatjes en films die mensen jarenlang op hun computer bewaren. Erwin: "Er wordt slecht weggegooid. Als politie een huis van een verdachte binnengaat, staan er overal pc's: bij de kinderen op de kamer, in de kelder." Ook die moeten worden onderzocht op digitale sporen – gemiddeld zo'n tien miljoen per zaak. Ga er maar aan staan.

Petabytes

Dus ontwikkelde het NFI een aantal jaar geleden Xiraf – kort samengevat: een forensische zoekmachine. Wie naar bewijs speurt, voert gestapeld filters in (alleen e-mails, alleen tussen datum x en y, alleen plaatjes) en beperkt zo de hoeveelheid data. De zoekmachine is zo eenvoudig, dat rechercheurs er zelf mee aan de slag kunnen en snel antwoord op hun vragen krijgen. Maar: Xiraf loopt nu op haar laatste benen. Reden voor Harm en Erwin om zich met vijf collega's op te sluiten in 'de garage' – een klaslokaal in het Field Lab waar drie dagen per week, zeven man hard zitten te werken aan een opvolger. Nodig, want elk jaar, berekenden Erwin en Harm conservatief, komt er zo'n 20 petabyte aan informatie binnen; goed voor 120 moderne computers per dag. De hoeveelheid data neemt alleen maar toe – en dat is problematisch.

"Xiraf gaat bij heel grote bestanden stuk", vat Erwin het probleem van de groeiende hoeveelheid data samen. "Als Xiraf een bestelbus is met een lading data, dan zakt dat busje straks door zijn assen. Dat punt is in 2014 ongeveer bereikt." Dus moest er een snelle oplossing komen. Harm: "Dan kan je een beetje sleutelen aan het bestaande systeem, wat er bij, wat er af, grotere laadbak, imperiaaltje erop - om Erwins voorbeeld door te trekken. Wij hebben echter gekozen voor een fundamentele oplossing: een nieuwe vrachtwagen met oplader." En die heet Hansken, genoemd naar de eerste olifant die in de zeventiende eeuw vanuit Sumatra in Nederland terecht kwam en door Rembrandt is getekend. Erwin: "Een vriendelijk savannedier, net als de giraffe, die als het moet overal doorheen gaat."

Wij-gevoel

De manier waarop Hansken wordt gebouwd, is uniek. Nooit eerder zonderde een groep NFI-medewerkers zich bewust af van het dagelijks werk om samen, dicht op elkaar, aan één project te werken. Een ideetje van Harm en Erwin zelf. "En het werkt", zegt Erwin. De lijnen zijn kort, problemen worden sneller getackeld, er gaat nauwelijks tijd verloren aan coördinatie van taken omdat iedereen in dezelfde ruimte op elkaars lip zit en precies weet waar de ander mee bezig is. Problemen en oplossingen vliegen door de ruimte en worden direct omgezet in praktische toepassingen. "Er heerst hier een sterk 'wij-gevoel': we zijn iets heel gaafs aan het maken dat met Kerst af moet zijn", zegt Erwin. Een strakke deadline en dan kost het overigens nog een jaar om Hansken operationeel te laten zijn.

Nihil

Wat gaat de rechercheur van de nieuwe Xiraf merken? Harm: "Weinig." De interface – het scherm waarmee rechercheurs hun data met Hansken onderzoeken, blijft hetzelfde. Achter dat digitale dashboard zit straks wel een heel nieuw systeem. Daarom zijn rechercheurs ook nauwelijks betrokken bij de ontwikkeling van Hansken en komen zij later in het proces aan bod. "De 'bottleneck' is de tijd die een opdracht duurt. Als we te lang op een resultaat moeten wachten, heeft het geen tactische en sturende waarde meer voor een onderzoek", legt Harm uit.

'Hansken is een vriendelijk savannedier die als het moet overal doorheen gaat'

De data van Hansken staat straks op particuliere servers en dat is voor velen binnen het NFI even wennen. Harm: "Veiligheid staat bovenaan. Als we moeten kiezen tussen veiligheid en, zeg: efficiëntie, dan kiezen we voor veiligheid." Om gebruikers daarvan te overtuigen, zijn alle principes waar het team van uit is gegaan, straks gewoon openbaar. Erwin: "Dat is ook bij openbare encrypties op internet zo: we weten precies hoe het principe werkt. Dat betekent niet dat ze de codes kunnen kraken." En zo is het straks ook bij Hansken: iedereen mag weten hoe de veiligheid is gewaarborgd, de kans dat je met die kennis het systeem binnenkomt, is nihil. Harm: "We hebben geen huis gebouwd en zetten daar vervolgens een hek om. Veiligheid is vanaf dag één onderdeel van het ontwerp geweest: *security by design*." Wie niets met de zaak te maken heeft, komt niet bij de data.
Rest de vraag of dat alleen maar in een 'garage' kan worden bedacht? "De grootste internet- en computerbedrijven zijn in een garage begonnen", lacht Erwin. Op deze plek maakt het team onverwachte zaken mee. "Laatst hadden we hier de helikopter voor Prinsjesdag op het platform staan. Toen klapperden de deuren." <<